Class II HLA interactions modulate genetic risk for multiple sclerosis

The International Multiple Sclerosis Genetics Consortium

Supplementary Information

Table of Contents

1 Data

The data consists of 11 country-level cohorts of European ancestry: Australia and New Zealand (AUSNZ), Belgium (BEL), Denmark(DEN), Finland (FIN), France (FRA), Germany (GER), Italy (ITA), Norway (NOR), Sweden (SWE), the United Kingdom (UK) and the United States of America (USA). In total there are 47,850 individuals, and the per population breakdown of the sample sizes, after quality control (QC) and exclusions, can be found in Supplementary Table 1. Full details of the QC that was performed can be found in the Supplementary Materials of the first multiple sclerosis Immunochip analysis[1].

Briefly, sample-level quality checks include gender checks (samples excluded if reported gender was inconsistent with the observed based on sex chromosome markers), call rate (samples excluded if <98%), autosomal heterozygosity (samples excluded if more than 3 standard deviations from the mean), ambiguity or inconsistency in the Sequenom finger print ID, and an excess of IBD sharing (PI_HAT>0.25). Contrary to the previous analysis[1], here we have included individuals that overlapped with the 2011 GWAS study[2].

SNP-level quality control was carried out for each population separately, using individuals that passed sample QC. The QC consisted of checking (in the following order) whether the SNPs had a call rate of <98%, had a HWE P value < 10^{-5}, exhibited differential missingness between cases and controls with a P value < 10^{-3} or were monomorphic. Only SNPs that passed QC in all populations were kept.

Principal components (PCs) were also calculated separately for each population, using the samples and an LD pruned set of SNPs that passed QC. An additional 1330 samples were removed as being outliers from the principal component analysis (PCA). Further details on the choice of SNPs for the PCA calculations can be found in Ref. S1. Full Sample and SNP exclusion lists by cohort can be found in Suppl. Tables 10 and 11 of Ref. S1, respectively.

In total, in this study we consider 6,218 SNPs from within the extended MHC region[3], from 29.9Mb to 33.6Mb (Hg19 / GRC37) on Chromosome 6.

In addition to disease status, we also consider the following secondary phenotypes:

- Multiple Sclerosis Severity Score (MSSS).
- Age at onset (AAO).
- Clinical course (primary progressive or relapsing-remitting disease progression).

Details on the availability of the secondary phenotypes across the samples can be found in Supplementary Table 5.

2 Imputation of HLA alleles

Classical HLA types were imputed using the program HLA*IMP:02[4]. Briefly, HLA*IMP:02 uses a graphical representation of the haplotype structure ("haplotype graph") in the extended MHC region to statistically infer a sample's classical HLA alleles. For this study, we created a haplotype graph from a pan-European reference panel ("GS&HLARES_EU" from Ref. S5) combining sample data from the CEU[3], CEU+[3], 1958 Birth Cohort (http://www.b58cgene.sgul.ac.uk) and HLARES[4,5] datasets.

GS&HLARES_EU comprises 6056 SNPs across the xMHC (as defined in Ref. S5) and a varying number of individuals with classical HLA type data available (1864 individuals with at least one 4-digit allele defined for *HLA-A*; 2630 for *HLA-B*; 1502 for *HLA-C*; 366 for *HLA-DQA1*; 2031 for *HLA-DQB1*; 2414 for *HLA-DRB1*; 74 for *HLA-DPB1*; 282 individuals with at least one 2-digit allele available for *HLA-DRB3*; 282 for *HLA-DRB4*; 282 for *HLA-DRB5*).

In the process of standard-HLA*IMP:02 pre-imputation quality control, all SNPs with more than 20% "missing data" in the MS dataset were removed and for complementary SNPs strandedness was aligned to HapMap. These steps were carried out with the standard HLA*IMP front-end and are described in detail in the Supplementary Information of the 2011 GWAS[2]. It should be noted that these QC steps were conducted on the already QC-filtered dataset that was used for the genome-wide analysis of MS Immunochip data.

HLA alleles were imputed at a 4-digit resolution for loci *HLA-A*, *HLA-C*, *HLA-B*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1* and *HLA-DPB1*, and at a 2-digit resolution for loci *HLA-DRB3*, *HLA-DRB4* and *HLA-DRB5* as no 4-digit resolution data were available as reference for the latter.

3 Validation of classical allele imputation for reported alleles

We assessed imputation accuracy by conducting a 2/3 (training) – 1/3 (validation) cross-validation experiment on samples of known HLA types, used as a reference panel for the imputation. We employed the same statistical model and the identical set of informative SNPs as were employed in the analysis. At 4-digit resolution, without imposing a threshold for calling on the posterior probabilities (call rate=100%), accuracy ranged from 0.9 to 0.99 (See Supplementary Table 2).

We also investigated the sensitivity, specificity and positive predictive value (PPV) of all alleles found to be associated with MS or secondary phenotypes (See Supplementary Table 3). Specificity was extremely high across all putatively associated alleles. Sensitivity and PPV were above 90% for all alleles found to be associated, apart from *B*38:01* (PPV = 0.89, *DRB1*08:01* (PPV = 0.84) and *DRB1*01:01* (Sensitivity = 0.82, PPV = 0.90). No systematic pattern in mis-imputation was found for *B*38:01*, while the majority of mis-imputations for *DRB1*01:01* and *DRB1*08:01* were with other alleles of the same 2-digit supertype (*e.g. DRB1*01:03* and *DRB1*08:02*, respectively).

4 Building a model of HLA risk for multiple sclerosis

Our aim was to identify the key classical HLA drivers of genetic risk for multiple sclerosis (MS), and other effects within the region that are of comparable significance but are not explained through linkage disequilibrium to specific HLA alleles. We use three different approaches to build such a model, taking a consensus strategy to summarise findings.

4.1 Approach 1: Manually curated search with UK focus

Approach 1 focused initial discovery on the UK cohort, which is the largest of all and most closely matched to the imputation training data, and then combine evidence across all cohorts using fixed effects meta-analysis (FEM) to confirm (validate and replicate) the association, estimate effect size (odds-ratios) and test for potential heterogeneity between cohorts. We recognise that if HLA allele imputation were perfect and population stratification were completed controlled this approach would

lose power relative to a simple joint analysis. Our approach, therefore, is conservative, but robust to particular failures in imputation or control of population stratification.

We started by incorporating in our model the four HLA alleles which were found to be significantly associated with multiple sclerosis in the 2011 GWAS[2], namely *DRB1*15:01*, *DRB1*13:03*, *DRB1*03:01* and *A*02:01*. As anticipated these all replicated in the FEM. Given that our training data only allowed 2-digit resolution of *DRB5* alleles it was not possible for us to fully resolve the relative contributions of the tightly correlated *DRB5*01:01* and *DRB1*15:01* alleles. Based on the known linkage disequilibrium of virtually 100% between *DRB1*15:01* and *DRB5*01:01* in populations with Caucasian ethnicity, the relative contribution of either allele to MS cannot be addressed in this study.

We then employed a stepwise logistic regression approach where additional parameters (both HLA alleles and SNPs in the MHC region) were considered for addition to the model if the strength of association in the relevant conditional analysis within the UK cohort had $P < 10^{-5}$. We only report effects which had a combined $P < 10^{-9}$ in the FEM of the model containing all other alleles identified through this iterative process (the 'full' model). 5 PCs were included as covariates in all models (these PCs were calculated from SNPs genome-wide, not including any from the xMHC region, as described in the 2013 multiple sclerosis Immunochip study[1]). Gender was not included as a covariate. Where appropriate, we employed likelihood ratio tests (LRT) to choose amongst competing models where they were nested, and otherwise the Bayesian Information Criterion[6] (BIC). Below we give details of the models considered and the steps taken at each stage of the selection procedure. A summary of the factors identified at each step is shown in Supplementary Figure 2.

To check consistency, we also rebuilt the model without any prior inclusion. All the alleles and effects reported above and below were included, though in a slightly different order. In particular, the risk effect of *DRB1*03:01* appears later (though is still included).

4.1.1 Baseline model with *DRB1*15:01, DRB1*03:01, DRB1*13:03, A*02:01* and departures from additivity

We started by assessing the evidence for deviations from additivity (on the log-odds – logistic - scale) for all four alleles of our baseline model, focusing primarily on the UK cohort. Specifically, we compared a baseline logistic model where the additive term is:

$$M_{ADD} = b_0 + G_{HLA} * b_1 + PC1 * b_{PC1} + PC2 * b_{PC2} + PC3 * b_{PC3} + PC4 * b_{PC4} + PC5 * b_{PC5},$$

where $G_{HLA} \epsilon \{0,1,2\}$ is the genotype for the respective HLA allele, to a model with an additional parameter which captures deviations from additivity for that HLA allele (in blue), coded as a correction of the effect for samples homozygous for the respective allele:

$$M_{IND} = b_0 + G_{HLA} * b_1 + I_{HLA\_HOM} * b_{1_{HOM}} + PC1 * b_{PC1} + PC2 * b_{PC2} + PC3 * b_{PC3} + PC4 * b_{PC4} + PC5 * b_{PC5},$$

where $I_{HLA\_HOM} = \begin{cases} 1, if\ G_{HLA} == 2 \\ 0, otherwise \end{cases}$. We refer to this parameter as the *homozygote correction term* for the allele. Models M_{ADD} and M_{IND} are nested, and the P value for the added term is obtained from the LRT. Each allele was assessed in turn.

For *DRB1*15:01*, there was significant evidence for departures from additivity, with $P = 7 \times 10^{-6}$ for the homozygote correction term and a negative $b_{1_{HOM}}$, indicating that the effect of homozygotes is

significantly less than double that of the heterozygotes in the log-scale (i.e. the risk effect is partially dominant). The effect remained strong and consistent in the 4-allele model (including parameters for *DRB1*03:01*, *DRB1*13:03* and *A*02:01*). It was therefore decided to include two parameters for the *DRB1*15:01* effect in the model.

For *DRB1*03:01*, the homozygote correction term had much stronger evidence for association with the trait in the UK cohort than the additive term ($P = 3x10^{-11}$ versus $P = 4x10^{-3}$, respectively). Moreover, the additive term appeared to have a protective effect on the trait in the single-allele model (only *DRB1*03:01* included) but a deleterious one in the 4-allele model. We therefore only include a single parameter for *DRB1*03:01*, assuming a recessive effect for that allele.

For *DRB1*13:03*, there was no evidence for departures from additivity ($P > 0.05$ for the homozygote correction term) in neither the single-allele nor the 4-allele models. Therefore, a single parameter for this allele was added in the baseline model, assuming an additive effect on the log-odds scale.

For *A*02:01*, the homozygote correction term had a negative coefficient, indicating that the effect of two copies of the allele on disease was less than double that of a single allele, but was only marginally significant ($P = 0.043$) in the UK cohort. In the 4-allele model, the effect was consistent with that in the single allele model, but with $P = 0.08$). Nevertheless, in the light of the interaction analysis (see below) showing that deviations from additivity in *A*02:01* can mimic weak interactions with *A*02:01* and non HLA loci (Supplementary Fig. 6), it was decided to include two parameters for *A*02:01* in the model.

The resulting linear term of in the baseline model is:

$$M_4 = b_0 + G_{DRB1501} * b_1 + I_{DRB1501HOM} * b_{1HOM} + G_{A0201} * b_2 + I_{A0201HOM} * b_{2HOM} + G_{DRB1303} * b_3 + I_{DRB0301HOM} * b_{4HOM} + PC1 * b_{PC1} + PC2 * b_{PC2} + PC3 * b_{PC3} + PC4 * b_{PC4} + PC5 * b_{PC5}$$

All terms replicate and validate in the FEM.

4.1.2 The signal of association at *DPB1*

When all HLA alleles and SNPs in the region were added in turn in to the above model (assuming an additive effect in the log-scale), the strongest subsequent signal in the UK is seen at rs9277565 with $P = 1x10^{-22}$ (compared to the most associated classical HLA allele, *DPB1*03:01*, with $P = 5x10^{-16}$). rs9277565 is in moderate LD with *DPB*03:01* (r^2 of 0.4 to 0.7 across cohorts) and it is likely that they are capturing the same effect (no *DPB1* signal is seen after conditioning on the SNP). Given the higher strength of association it was decided to include rs9277565 in the model rather than the DPB allele. There was no evidence for departure from additivity ($P > 0.05$ for the homozygote correction term). The effect replicates and validates in the FEM.

4.1.3 A protective effect for *B*44:02*

In the 6-allele model, across all SNPs and classical HLA alleles, the most significant effect in the UK was for *B*44:02* ($P = 1x10^{-12}$) and *B*44:02* was therefore also included in the model; assuming an additive effect on the log-odds scale since there was no evidence for departures from additivity ($P > 0.05$). The effect replicates and validates in the FEM. We note that *B*44:02* is in moderate LD with *C*05:01* ($r^2 = 0.56$ in controls), and likely explains previous reports of association of this allele with MS risk[7].

4.1.4 A protective effect of the rare allele *B*38:01*

Across 7-allele models, the strongest signal in the UK came from a SNP, rs2229092 ($P = 8x10^{-11}$). Among classical HLA alleles, the strongest signals were observed at *C*15:02* ($P = 2x10^{-7}$) and *B*38:01* ($P = 7x10^{-7}$). However, *C*15:02* was found to be in high LD with SNPs which showed stronger evidence for association with the disease (e.g. $r^2 = 0.86$ with rs16899166 and $r^2 = 0.91$ with rs5009853). *B*38:01* on the other hand, showed stronger evidence for association in comparison with all SNPs it was in LD with (and showed no association, $r^2 > 0.01$, to any SNP with stronger association). We therefore included *B*38:01* in the model at this stage. No evidence for departures from additivity was detected. The effect replicates and validates in the FEM.

4.1.5 A risk effect for *DRB1*08:01*

In the 8-allele model, *C*15:02* ($P = 3x10^{-7}$) and rs2229092 ($P = 1x10^{-10}$) remained the most associated HLA allele and SNP in the UK, respectively. The second most strongly associated HLA allele was *DRB1*08:01* ($P = 2x10^{-6}$). Notably this allele had shown suggestive evidence for association in the FEM analysis of the original 2011 GWAS[2] ($P = 2x10^{-7}$), but was not highlighted as associated in that study as it showed only marginal evidence of association in the UK cohort alone ($P = 0.025$). The same allele was also reported previously by another study, though also not at genome-wide significance[8]. Given this prior evidence for association of *DRB1*08:01* with MS ,the fact that it was not in high LD ($r^2 > 0.8$) with SNPs with stronger evidence for association, and shows no correlation with *C*15:02* or rs2229092 ($r^2 < 0.01$), we included *DRB1*08:01* in the model. The effect replicates and validates in the FEM. The other effects are discussed below.

4.1.6 Interaction between *DRB1*15:01* and *DQA1*01:01*, and an effect associated with *LTA*

Based on our interaction analysis we next considered an *DQA1*01:01-DRB1*15:01* interaction term (i.e. a parameter indicating an effect of *DQA1*01:01* only when at least one copy of *DRB1*15:01* allele is carried by the individual*;* see section below on searching for HLA-HLA interactions). The inclusion of this interaction term leads to a better fit than all models containing any of the non-interaction variants ($P = 2x10^{-12}$). The interaction term replicates in the FEM and was therefore included in the model. We also note that the interaction term shows strong association in a model with *DRB1*15:01* alone ($P = 3x10^{-12}$ with 2 parameters for DRB1*15:01). In contrast, the P-value for a marginal effect of *DQA1*01:01* is 0.033. In the full model, the marginal effect for *DQA1*01:01* has $P > 0.1$, so both before and after the other terms have been included in the model, the interaction term shows substantially stronger evidence for association than a marginal *DQA1*01:01* term.

When including the interaction term in the model, the signals at *C*15:02* and rs2229092 remain. To further evaluate their roles, we ran the FEM on models including the 9 established parameters (2 for *DRB1*15:01*, 2 for *A*02:01*, a recessive effect for *DRB1*03:01*, additive effects for *DRB1*13:03*, *DRB1*08:01*, *B*44:02*, *B*38:01*, rs9277565, and the *DQA1*01:01-DRB1*15:01* interaction term) together with each of *C*15:02* and rs2229092 in turn. In the analysis including rs2229092, this SNP exhibited strong evidence for association ($P = 1x10^{-19}$) in the FEM. In contrast, in the analysis including *C*15:02*, the evidence for association failed to reach the 10^{-9} threshold ($P = 6x10^{-9}$) and is seen to be much weaker in the other non-UK cohorts (FEM $P = 0.0015$). This *C*15:02* allele is in high LD with SNPs exhibiting stronger signals of association throughout our stepwise regression approach. Consequently, while our results are supportive of a possible role for *C*15:02* in risk for MS, the level of evidence does not meet the standards applied here for inclusion. We also considered models where rs2229092 was included at earlier stages of the model and found no material changes in the results for other reported effects. Similarly, when including only rs2229092 (together with 5 PCs as

covariates), it is strongly associated in the UK ($P = 7x10^{-5}$). We therefore included rs2229092 in the model rather than *C*15:02*.

It should also be noted that while rs2229092 (a missense variant in the *LTA* gene) exhibits the strongest signal in the 10-allele model in the UK and is highly significant in the FEM, it is moderately correlated (r^2~0.45) with other SNPs which show strong evidence for association, such as rs225798 and rs2230365, both in the *NFKBIL1* gene; these alternate SNPs are also candidates for driving this observed association. While the possibility that rs2229092 (or another of these SNPs) are tagging one (or more) signals coming from HLA alleles cannot be ruled out, the evidence for association is substantially stronger for the SNP than for any HLA alleles included in our analysis suggesting this is not the case, at least for more common HLA alleles. We note that rs2229092 is in moderate LD (r^2 = 0.33 in controls) with the previously reported rs2516489[9] (and generally lies within an extended LD block), though the latter shows weaker association in this analysis.

4.1.7 A deleterious effect for DQB1*03:02 and evidence for interaction with DQB1*03:01

When looking across the other cohorts for HLA alleles with remaining evidence for association in the 11-allele model, the only allele with $P < 1x10^{-5}$ in any of the cohorts is *DQB1*03:02* in the SWE cohort ($P = 2x10^{-8}$). *DQB1*03:02* exerts a dominant effect on the log-odds scale and also appears significantly associated with MS in the FEM across cohorts ($P = 2x10^{-13}$). Moreover, there is supporting evidence for this effect coming from the automated approaches (see sections 4.2, 4.3 of the supplement). Therefore, it was decided to include this allele in the model as well. Interestingly, when conducting the interaction analysis (see section 7 of the supplement for details) for DQB1*03:02, we found strong evidence ($P<10^{-6}$) for an interaction between DQB1*03:02 and DQB1*03:01 in the UK cohort (Supplementary Figure 3). When including the interaction term in the full model, the effect of the interaction term was significantly associated with MS in the FEM across cohorts ($P = 10^{-14}$). Moreover, the DQB1*03:02 association with MS became somewhat stronger and even more significant ($P = 4x10^{-24}$, compared to $P=2x10^{-13}$ before the interaction term was added, Supplementary Table Y).

4.1.8 Remaining signals in the UK and other cohorts

The only other HLA allele included in the model is B*55:01. Despite showing no evidence for association with MS with $P < 10^{-5}$ in any single cohort, it appeared significantly associated with MS in the automated approaches (see sections 4.2, 4.3 of the supplement) and its signal was validated in the FEM across cohorts for the full model ($P = 10^{-11}$). It confers a protective effect with no evidence for deviations from additivity.

Conditioning on the effects described above, no classical HLA allele showed evidence for association with MS with $P < 10^{-5}$ within the UK cohort in the full, 13-allele model. However, 14 SNPs did show evidence for association at that level, with the top hit, rs9267482 ($P = 7x10^{-10}$) lying in the *DDX39B* (*BAT1*) gene. This SNP is also highly significant in the FEM ($P = 10^{-21}$).

4.2 Approach 2: Automated model search at of 4-digit HLA alleles with cross-cohort mega-analysis

To approach the construction of a general model for HLA-allele risk for multiple sclerosis in a manner that is not guided by prior knowledge, we used an automated model search strategy in which all cohorts are considered jointly (referred to as a mega-analysis). Prior to analyses we took the following steps:

- Setting the threshold for imputation to be 0 – i.e. taking the allele call with the highest posterior probability in each case and treating this as fixed. This step is needed to enable appropriate model comparison between loci. By way of comparison, setting a threshold of 0.5 would remove 2.2% of imputed genotypes across loci.
- Removing alleles with a combined frequency after imputation across cohorts of 0.5%. This step removed 55 alleles out of 232 across the loci considered.
- Removing alleles imputed to be in perfect association ($r^2 = 1$) with each other. This step removed 8 of 177 alleles.
- Removing individuals where the imputed allele was not present in the IMGT database (see below). This step removed 530/47,849 individuals.

Starting with a base-line model including effects for each cohort (as a factor) and principal components, at each stage we performed logistic regression on disease risk for every classical HLA allele in turn, considering a general genotype model (i.e. separate coefficients for each genotype). Consequently, the allele that led to the highest increase in likelihood was identified and in a separate step, the Bayesian Information Criterion (BIC) was used to decide whether the effect was best described as additive, recessive, dominant or general. A range of strategies for backwards elimination were considered, though in practice no allele was ever removed. The model selection process was run until the BIC no longer increased. However, only those effects that also achieve $P < 10^{-9}$ in the FEM are reported. A summary of the factors identified at each step and how these relate to the factors identified by the other approaches is shown in Supplementary Figure 2.

We note that we also considered an automated search including SNP variants. However, because most classical HLA alleles are well tagged by at least one (and often many) SNPs within the region, we commonly observed that effects that are typically interpreted as being driven by classical HLA alleles were assigned to SNPs (potentially due to errors in imputation and / or chance fluctuations in association). We therefore only considered classical HLA alleles in the automated approach, though compared results to the manually curated selection process to check consistency.

4.3 Approach 3: Automated model search augmented with allele groupings at 2-digit level and by sharing of amino acids at variable residues

Previous research has demonstrated that some association between groups of classical HLA alleles and genetic risk for disease can be explained by the sharing of particular amino acid residues at variable sites within the mature protein[10]. We therefore considered a separate automated model search strategy on the combined cohorts in which we augmented the set of HLA alleles by genotypes at allele groups defined by both 2-digit resolution and by the sharing of specific amino acid residues at variable sites. Group membership was inferred from the imputed allele at 4-digit resolution. Aligned amino acid sequences for the imputed alleles were obtained from IMGT (http://www.ebi.ac.uk/ipd/imgt/hla/align.html) on March 4th 2015. In a few instances, full amino acid sequences were not available for all alleles imputed.

As above, we took the following steps:

- Setting the threshold for imputation to be 0 – i.e. taking the allele call with the highest posterior probability in each case and treating this as fixed.
- Removing alleles / allele groups with a combined frequency after imputation across cohorts of 0.5%. This step removed 108 alleles / allele groups out of 454 across the loci considered.
- Removing alleles / allele groups imputed to be in perfect association ($r^2 = 1$) with each other. This step removed 22 of 346 alleles / allele groups.
- Removing individuals where the imputed allele was not present in the IMGT database. This step removed 530/47,849 individuals.

As above, we considered a general genotype model for each allele / allele group at each step, starting from a baseline model with coefficients for each cohort and PCs. At each stage the allele or allele group leading to the greatest increase in likelihood was identified and BIC used to assign additive, dominant, recessive or general models. The procedure was run until BIC no longer increased, though only effects that achieve $P < 10^{-9}$ in the FEM are reported. A summary of the effects identified at each stage is shown in Supplementary Fig. 2.

4.4 Consensus strategy and comparison

The three model selection strategies identified a series of related, though non-identical effects (Supplementary Fig. 2). Moreover, in two cases, effects identified by the automated approaches were better described as interactions. For example, the homozygous protective effect identified as associated with DQB1 alleles with a glycine at residue 70 in step 7 of the augmented model selection process is better explained (the data is 400 times more likely) through an interaction with *DRB1*15:01.*

We therefore used a consensus strategy to combine results from the different approaches to give an overall picture of genetic risk for multiple sclerosis, using linkage disequilibrium between risk factors to identify sets of related factors (Supplementary Figure 2). Each effect is labelled with the most associated 4-digit allele, though we acknowledge the subjectivity in this choice.

5 Validation of top hits by comparison of uncertainty-aware and uncertainty-unaware analyses

In order to ensure that the identified associations were not affected by mis-imputation, additional analyses where conducted where the uncertainty associated with imputation (quantified by the posterior probability of each call) was incorporated into the logistic regression framework, with numerical optimisation used to find maximum likelihood estimates. Specifically, models of the following form were run:

$$M = b_0 + G_{DRB1501} * b_1 + I_{DRB1501HOM} * b_2 + G_{A0201} * b_3 + I_{A0201HOM} * b_4 + G_{OTHER} * b_{OTHER} + PC1 * b_{PC1} + PC2 * b_{PC2} + PC3 * b_{PC3} + PC4 * b_{PC4} + PC5 * b_{PC5}$$

where two parameters were included for each of *DRB1*15:01* and *A*02:01*, as well as a parameter for each other associated allele in turn, together with 5 PCs. Specifically, $G_x \epsilon \{0,1,2\}$ is the genotype for the respective HLA locus x, and I_x a correction to the additive term for homozygotes, $I_x = \begin{cases} 1, if\ G_x == 2 \\ 0, otherwise \end{cases}$. The effect assumed for all other susceptibility alleles was the same as in the

uncertainty-unaware models (additive in the log-scale for all but *DRB1*03:01* which was modelled as purely recessive).

The results obtained from the above model when applying uncertainty-aware analysis were qualitatively identical to these obtained from the uncertainty-unaware one (with 100% call rate), in terms of both effect size and P value for significance. These are shown across cohorts for all associated alleles in Supplementary Fig. 1.

6 Secondary phenotypes

Secondary phenotypes, such as clinical course (Primary Progressive vs Relapsing Remitting MS), severity (calculated as Multiple Sclerosis Severity Score, MSSS, which reflects the rate at which patients affected with the disease accumulate disability[11]) and age at onset were available for a proportion of the samples studied (Supplementary Table 5). At first, we sought to assess the evidence for an effect of the combined HLA risk score on each of the secondary phenotypes. The HLA risk score was estimated from the effect sizes of the alleles included in the 10-allele model and reported in the main text, as they were estimated from the FEM, as:

$$HLA_{RISK\ SCORE} = \exp\left(G_{DRB1501} * b_1 + I_{DRB1501HOM} * b_{1HOM} + G_{A0201} * b_2 + I_{A0201HOM} * b_{2HOM} + G_{DRB1303} * b_3 + I_{DRB0301HOM} * b_{4HOM} + G_{DRB0801} * b_5 + G_{B4402} * b_6 + G_{B3801} * b_7 + G_{rs9277565} * b_8 + G_{DQA0101} * I_{DRB1501} b_9 + G_{rs2229092} * b_{10}\right)$$

where $G_x \epsilon \{0,1,2\}$, $I_x = \begin{cases} 1, if\ G_x == 1 \\ 0, otherwise \end{cases}$, $I_{xHOM} = \begin{cases} 1, if\ G_x == 2 \\ 0, otherwise \end{cases}$ for each HLA allele or SNP x.

The HLA risk score was quantile normalised within each cohort. We used the HLA risk score as the explanatory variable and conducted the following analyses:

- linear regression with AAO (which is approximately normally distributed)
- linear regression analysis with MSSS, using the quantile-normalised MSSS values as the phenotype,
- logistic regression analysis of extremes on MSSS, splitting the patients into two groups of low (MSSS<2.5) and high (MSSS>7.5) severity and using the resulting binary grouping as the phenotype value, and
- logistic regression analysis on clinical course (PPMS vs RRMS).

For all models, we included 5 PCs as covariates. When running these on the UK cohort, the HLA risk score was found nominally associated with AAO (P = 0.042) but not with MSSS or the clinical course. The association of HLA risk score with AAO was confirmed in a FEM across cohorts ($P = 7x10^{-10}$, Fig. 4A). Specifically, the difference in age of onset between individuals lying at the 5th and 95th percentiles is approximately 2 years (34.23 and 32.26 years respectively).

6.1 Age at onset

In order to assess the effect of specific HLA alleles on the age at onset (AAO) of MS, we conducted linear regression analysis in a stepwise fashion, considering AAO as the phenotype and including 5 PCs as covariates in all models. All alleles were assumed to exert an additive effect on the trait. The models were run separately for each cohort and results were subsequently combined by running a

fixed-effects meta-analysis for alleles with nominally significant associations (and the same direction of effect) in two or more cohorts. As for disease risk, joint analysis of the combined cohorts controlling from cohort effects and PCs was also carried out using an automated model search procedures with and without allele groupings based on 2-digit classifications and sharing of amino acid residues.

6.1.1 Manually curated approach with UK focus

From the manually-curated approach, across all single-allele models, the strongest signal came from *DRB1*01:01* ($P = 4x10^{-11}$, effect size = 1.82 years per allele). *DRB1*01:01* is in moderate LD with *DQA1*01:01* ($r^2 = 0.6$) and *DQB*05:01* ($r^2 = 0.7$), which also showed comparable effects in the same direction. The second strongest signal came from *DQB1*06:02*, which decreases the age at onset of MS ($P = 6x10^{-11}$, effect size =-0.98 years per allele). *DRB1*15:01* and *DQA1*01:02*, are both part of the extended MS risk haplotype containing *DQB1*06:02*, and thus as expected also exhibited strong associations in the same direction (Supplementary Fig. 7).

After conditioning on *DRB1*01:01*, *DQB1*06:02* was the only allele that remained associated with AAO of MS at the genome-wide significance threshold ($P = 2x10^{-8}$, effect size = -0.76 years per allele), with *DRB1*15:01* narrowly missing this ($P = 8x10^{-8}$, effect size= -0.72 years per allele). With these data it is not possible to definitively establish which of these class II alleles is driving the observed association. Hence, and because *DRB1*15:01* is driving the association with the main phenotype, it was decided to include *DRB1*15:01* in the model. When *DRB1*15:01* was also included in the model, no other allele showed genome-wide significant evidence for association with AAO.

In order to ensure that the effect of *DRB1*01:01* on AAO is not tagging an effect from *DQA*01:01* in the presence of *DRB1*15:01* or *vice versa,* we conducted a formal model comparison between models with

- both *DRB1*01:01* and the interaction term included ('BOTH'),
- only *DRB1*01:01* included ('DRB0101 ONLY'), or
- only the interaction term included ('DQA0101INT ONLY'),

for both the binary MS phenotype and the AAO. 5 PCs were included as covariates in both models. All other susceptibility alleles were also included in the respective models. Results from the UK cohort indicate that a model with only *DRB1*01:01* in fits the data better for AAO (both AIC and BIC), whereas a model with only the *DQA*01:01 - DRB1*15:01* interaction term in fits the data better for the main MS phenotype (both AIC and BIC); data not shown. In summary, the effects of the *DQA1*01:01 - DRB1*01:01* haplotype MS risk and AAO are best explained through distinct models but the evidence separating the alternative is not overwhelming.

6.1.2 Automated model search across combined cohorts

The automated model searches (with and without allele groupings) were carried out as for disease status on the subset of individuals for which AAO data was available. The primary factor identified when just 4-digit alleles were considered was, as with the manually-curated approach, an additive effect for *DQB1*06:02* (each copy reducing AAO by 1.2 years), with a subsequence general genotype effect for *DRB1*01:01* (Het genotype increase AAO by 3.9 years, Hom increases AAO by 5.2 years). When allele groupings were also considered, the primary effect identified (a general genotype model associated with sharing of a histidine residue at position 30 among DQB1 alleles) correlates with

*DQA1*01:01* (r = 0.67), *DRB1*01:01* (r = 0.52) and *DQB1*06:02* / *DRB1*15:01* (-0.37 / -0.32), thus incompletely capturing the two effects associated with classical HLA alleles.

6.2 Severity and clinical course

In order to further assess the effect of HLA alleles on the severity of MS, we more fully explored the relationship between HLA alleles and Multiple Sclerosis Severity Score (MSSS)[11]. We conducted

- linear regression analysis, using the quantile-normalised MSSS values as the phenotype and
- logistic regression analysis of extremes, splitting the patients into two groups of low (MSSS<2.5) and high (MSSS>7.5) severity and using the resulting binary grouping as the phenotype value.

For both sets of models, 5 PCs were included as covariates. Under both sets of models, no HLA alleles were found to be significantly associated with the severity of MS. There was some suggestive evidence for a modest effect of *B*08:01* (which is correlated with *DRB1*03:01*, r = 0.73) towards a more severe disease manifestation, more so for the binary analysis (FEM P = 2 x 10^{-5}, OR = 1.32). However, this does not reach the level of evidence applied throughout the analysis and is not analysed further.

In order to assess the effect of HLA alleles on the clinical course of MS (Primary Progressive vs Relapsing Remitting MS), we conducted logistic regression analysis, considering clinical course as the phenotype and including 5 PCs as covariates in all models. Using this model, we found no evidence for association between HLA alleles and clinical course of MS (No alleles with P > 0.05 in the UK and / or P > 10^{-5} in the FEM).

7 Interactions among classical HLA alleles

In order to investigate potential interactions between the HLA alleles included in our model and other HLA alleles, we first ran models with:

- two parameters (one additive, one homozygote correction term) included to model the effects of *DRB1*15:01* and *A*02:01*,
- a single parameter included for each other SNP/HLA allele identified as associated with MS in this analysis (additive on the log-odds scale for all alleles apart from *DRB1*03:01, DQB1*03:02*, where a recessive and a dominant effect, respectively, was assumed),
- a single parameter (additive on the log-odds scale, in blue below) for each other HLA allele in turn, and
- an interaction term (in bold, below) which models the effect of each other HLA allele in the presence of the allele under consideration (example below used for investigating interactions with *DBR1*15:01*).

Models have the following form:

$$M_{INT-DRB1501} = b_0 + G_{DRB1501} * b_1 + I_{DRB1501HOM} * b_2 + G_{A0201} * b_3 + \mathrm{I_{A0201HOM}} * \mathrm{b_4} + G_{DRB1303} * b_5 + I_{DRB0301HOM} * b_6 + G_{DRB0801} * b_7 + G_{B4402} * b_8 + G_{B3801} * b_9 + G_{B5501} * b_{10} + I_{DQB0302} * b_{11} + G_{rs9277565} * b_{12} + G_{rs2229092} * b_{13} + G_{otherHLA} * b_{14} + \boldsymbol{I_{DRB1501} * G_{otherHLA} * b_{15}} + PC1 * b_{PC1} + PC2 * b_{PC2} + PC3 * b_{PC3} + PC4 * b_{PC4} + PC5 * b_{PC5}\ .$$

For all models, $G_x \epsilon \{0,1,2\}$ is the genotype for the respective SNP/HLA allele, I_{xHOM} a homozygote correction term for HLA allele x, $I_{xHOM} = \begin{cases} 1, if\ G_x == 2 \\ 0, otherwise \end{cases}$, and $I_x = \begin{cases} 1, if\ G_x == 1 \\ 0, otherwise \end{cases}$ is an indicator of the presence of HLA allele x in an individual.

QQ plots of the interaction term for these analyses are shown in Fig 2 and Supplementary Fig. 3 for the UK cohort. There is evidence ($P < 10^{-6}$) for an interaction between *DRB1*15:01* and three strongly-associated HLA alleles: *DQA1*01:01* ($P = 3x10^{-9}$), *DQB*05:01* ($P = 4x10^{-8}$) and *DRB1*01:01* ($P = 1x10^{-7}$), all conferring a protective effect on the disease. The association is confirmed in the FEM across cohorts ($P = 2x10^{-19}$). These alleles (*DQA1*01:01*, *DQB*05:01* and *DRB1*01:01*) are all correlated (See Supplementary Table 4) and form part of a common European haplotype.

Moreover, there is evidence ($P=3x10^{-7}$) for an interaction between DQB1*03:02 and DQB1*03:01, again conferring a protective effect on the disease. This effect replicates in the FEM (Supplementary Table Y)

We also found modest evidence for interactions between *A*02:01* and other HLA alleles (though none with $P < 10^{-5}$ in the UK cohort) which was most pronounced for interaction with *DRB3*09*, *DRB5*01*, *DRB5*09* and *B*08:01*. However, when these interaction terms are included in turn in the 10-allele model, the fit is either better for the model without the interaction term (for *DRB3*09*, *B*08:01*) or the P value is not nominally significant in the UK cohort (for *DRB5*09*, *DRB5*01*). Therefore, there is no significant evidence in favour of interactions between *A*02:01* and other HLA alleles in MS.

With the exception of *B*44:02*, the QQ plots for other HLA allele show no deviation from the expectation under the null (Supplementary Fig. 3). For *B*44:02*, the inflated QQ plot shows a general increase in the P values of the interaction term, although no particular interaction stands out. This could indicate slight deviations from additivity, as seen for *A*02:01* prior to inclusion of a homozygote correction term for it in the model. However, the homozygote correction term for *B*44:02* was not nominally significant in either the single-allele or the full model.

8 Interactions between classical HLA alleles and non-HLA associated loci

We looked for evidence of interaction between HLA risk alleles and non-HLA associated MS risk loci identified in the recent Immunochip study[1], by adding a parameter for the effect of the non-HLA index SNP in the presence of each of the risk HLA alleles in turn in our model. Specifically, we started by looking at interactions between *DRB1*15:01* (and other alleles in turn) and the non HLA SNPs by fitting a model with 2 parameters for *DRB1*15:01*, an additive effect for the index SNP and an interaction term (in bold) between the presence of *DRB1*15:01* and the respective SNP (together with 5 PCs):

$$M_{INT-DRB1501} = b_0 + G_{DRB1501} * b_1 + I_{DRB1501HOM} * b_2 + G_{SNP} * b_3 + \boldsymbol{I_{DRB1501} * G_{SNP} * b_4} + PC1 * b_{PC1} + PC2 * b_{PC2} + PC3 * b_{PC3} + PC4 * b_{PC4} + PC5 * b_{PC5}$$

where $G_x \epsilon \{0,1,2\}$ is the genotype for the respective HLA allele or SNP x, $I_{DRB1501HOM}$ a correction to the additive term for DRB1*15:01 for homozygotes, $I_{DRB1501HOM} = \begin{cases} 1, if\ G_{DRB1501} == 2 \\ 0, otherwise \end{cases}$, and $I_{DRB1501} = \begin{cases} 1, if\ G_{DRB1501} == 1 \\ 0, otherwise \end{cases}$ is an indicator of the presence of *DRB1*15:01* in an individual.

We saw no evidence for interaction using this model (Fig 2 and Supplementary Fig. 4). For *A*02:01*, initial evidence for deviation from the expected uniform distribution of P values led to the identification of a weak non-additive effect of the allele, correction for which removes any evidence for HLA-nonHLA interactions (Supplementary Fig. 6). None of the alleles tested showed any departure from the null (Supplementary Fig. 4).

We also looked at whether the effect of HLA alleles on MS is stratified based on the cumulative risk from the non-HLA effects. To do this, we divided the samples (combined cases and controls within the UK cohort) into quartiles of non-HLA risk score (RS), defined by combining information across the L loci found associated in Ref S1. Specifically, the RS for an individual *j* was calculated as:

$$RS_j = \exp\left(\sum_{i=1}^{L} G_{ij} * b_i\right)$$

with $G_{ij} \epsilon \{0,1,2\}$ being the genotype for the respective non-HLA SNP i and i the corresponding log-odds ratio. As seen in Fig. 2 and Supplementary Fig. 5, there were no substantial differences in the RS-stratified odds-ratios compared to that estimated across all samples combined.

9 Estimating the contribution of polygenic epistasis

See Supplementary Note.

10 List of supplementary figures

Supplementary Figure 1. Effects of incorporating uncertainty in classical HLA allele prediction. Comparison of P values and odds ratios estimated for each cohort for all classical HLA allele effects shown in Fig. 1 under methods taking account of uncertainty in classical HLA allele imputation (y-axis) and not taking into account of uncertainty (x-axis). Values reported are from a model including additive and non-additive effects for *DRB1*15:01* and PCs, in addition to the indicated allele.

Supplementary Figure 2. Summary of the results from the model selection procedures (without interaction search) using manual curation (left), automatic model search with alleles at 4-digit resolution (middle) and automatic mega-analysis augmented with alleles grouped at 2-digit resolution and by sharing of amino acid residues. Alleles shown in order of inclusion in each approach and matched between approaches by linkage disequilibrium (Pearson correlation coefficient across entire cohort shown and groups identified by colours). At each stage, models with additive (no superscript), dominant (D), recessive (R) and general (G) risk were tested. The allele used to refer to each group is indicated by bold. Effects identified that are better explained by interactions between other alleles are underlined. Note that DRB1_AA57.S is a near-perfect proxy (r = 0.95) for the presence of either *DRB1*08:01* or *DRB1*13:03*, though a model with independent effects for the two alleles provides better fit. We note also that B_AA326.C is weakly correlated with the presence of *B*44:02* (r = 0.26), though the latter remains significant in FEM if the former is included.

Supplementary Figure 3. Evidence for interactions among classical HLA alleles affecting risk for multiple sclerosis. QQ plots showing the distribution of P values for the interaction term between HLA alleles reported in the full model and other classical HLA alleles. For each test, the full model, including non-

additive effects for *DRB*15:01*, *DRB1*03:01* and *A*02:01*, is fitted, along with the potential interaction.

Supplementary Figure 4. Evidence for interactions between classical HLA alleles and non-HLA disease-associated variants affecting risk for multiple sclerosis. QQ plots showing the distribution of P values for the interaction term between HLA alleles reported in the full model and non-HLA SNP variation influencing genetic risk for multiple sclerosis (from Ref. S1). For each test, the full model, including non-additive effects for *DRB*15:01*, *DRB1*03:01* and *A*02:01*, is fitted, along with the potential interaction.

Supplementary Figure 5. Interactions between HLA variants and combined non-HLA risk for multiple sclerosis. The effect of the indicated allele or variant among individuals stratified in to quartiles by a combined non-HLA risk score, obtained by multiplying odds-ratios associated with each genotype an individual carries at non-HLA loci influencing risk (from Ref. S1). The point estimate and 95% confidence intervals for effect size are estimated independently for each quartile of non-HLA genetic risk. Dashed and dotted lines indicate the combined point estimate and 95% confidence interval respectively. All analyses for the UK cohort only.

Supplementary Figure 6. Conflation between evidence for interactions and a departure from additivity for *A*02:01*. **A,** QQ-plot showing the distribution of P values for interaction terms between *A*02:01* and non-HLA variants known to affect risk of multiple sclerosis (from Ref. S1) under a model where *A*02:01* acts additively (on the log-odds scale). **B,** As for part A, but where a departure from additivity is included for *A*02:01*. Because potential interactions are modelled as an additional additive effect of second allele that only acts in the presence of at least one copy of *A*02:01*, such an effect can partly mimic additive and independent effects of the two loci, but with a departure from additivity at *A*02:01*.

Supplementary Figure 7. Classical HLA alleles affecting age-at-onset. Meta-analysis showing effects of pairs of classical HLA alleles in strong linkage disequilibrium (*DRB1*15:01 – DQB1*06:02* and *DRB1*01:01 – DQA1*01:01*) on age-at-onset.

11 List of supplementary tables

Supplementary Table 1. Number of cases and controls across the 11 cohorts investigated in the study.

Supplementary Table 2. Imputation accuracy at 4-digit type resolution in a 2/3-1/3 cross validation experiment without applying a call threshold on the posterior probabilities (call rate=100%).

Supplementary Table 3. Allele-specific sensitivity, specificity and positive predictive value (PPV) for alleles discussed in the main text. Statistics are calculated from 2/3 – 1/3 cross-validation experiment at a 4-digit level resolution, without applying a call threshold on posterior probabilities (call rate=100%).

Supplementary Table 4. HLA alleles and SNPs found to be significantly associated with multiple sclerosis in this study. All alleles/SNPs are presented in the order of inclusion in the model (see Supplementary Material above for details). Second column contains the number of parameters used for each allele in

the full model, while the third describes the effect assumed for each of the parameters (all effects are assumed in the log-odds scale). Columns 4-6 contain the P values, estimated odds ratios and 95% CIs, respectively, of the alleles / effects in the full model. Estimated values are from the fixed effect meta-analysis across cohorts. Columns 7-8 contain Cochrane's Q statistic for heterogeneity and the associated P value. Column 9 contains I^2, an estimate for the proportion of total variability explained by heterogeneity[12]. The last two columns contain the HLA alleles and SNPs, respectively, in LD with the alleles included in the full model (column 2). Entry "None" in these columns means that there is no allele with an r^2 of 0.5 or more with the reported allele. For HLA alleles (column 10), all alleles with an $r^2 > 0.5$ are reported. For SNPs (column 11), only the most associated ones are (with r^2 thresholds of 0.95, 0.9, 0.8 and 0.5).

Supplementary Table 5. Patient counts across cohorts for the secondary phenotypes analysed in this study.

12 Supplementary references

1. International Multiple Sclerosis Genetics Consortium. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet* **45**, 1353-1360 (2013).
2. The International Multiple Sclerosis Genetics Consortium & The Wellcome Trust Case Control Consortium 2. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214-219 (2011).
3. de Bakker, P.I. *et al.* A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet* **38**, 1166-1172 (2006).
4. Dilthey, A. *et al.* Multi-population classical HLA type imputation. *PLoS Comput Biol* **9**, e1002877 (2013).
5. Dilthey, A.T., Moutsianas, L., Leslie, S. & McVean, G. HLA*IMP--an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics* **27**, 968-972 (2011).
6. Schwarz, G.E. Estimating the dimension of a model. *Annals of Statistics* **6**, 461-464 (1978).
7. Yang, Y. *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med* **369**, 1502-1511 (2013).
8. Dyment, D.A. *et al.* Complex interactions among MHC haplotypes in multiple sclerosis: susceptibility and resistance. *Hum Mol Genet* **14**, 2019-2026 (2005).
9. Gonzaga-Jauregui, C., Lupski, J.R. & Gibbs, R.A. Human genome sequencing in health and disease. *Annu Rev Med* **63**, 35-61 (2012).
10. Patsopoulos, N.A. *et al.* Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet* **9**, e1003926 (2013).
11. Roxburgh, R.H. *et al.* Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity. *Neurology* **64**, 1144-1151 (2005).
12. Higgins, J.P.T. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* **21**, 1539-1558 (2002).

Supplementary Table 1. Number of cases and controls across the 11 cohorts investigated in this study.

Country	Abbr	Case	Control
Australia and New Zealand	AUSNZ	1021	947
Belgium	BEL	313	1705
Denmark	DEN	890	835
Finland	FIN	471	488
France	FRA	387	354
Germany	GER	2621	6308
Italy	ITA	962	1256
Norway	NOR	911	701
Sweden	SWE	2740	2872
UK	UK	4542	9359
USA	USA	2607	5560
ALL	ALL	17465	30385

Supplementary Table 2. Imputation accuracy at 4-digit type resolution in a 2/3-1/3 cross validation experiment without applying a call threshold on the posterior probabilities (call rate=100%, 2nd column), and with applying a threshold of T=0.7 (3rd column, with call rate shown at column 4).

Locus	Accuracy T=0	Accuracy T=0.7	Call Rate T=0.7
HLA-A	0.97	0.97	0.98
HLA-B	0.95	0.98	0.94
HLA-C	0.97	0.97	0.99
HLA-DPB1	0.9	0.98	0.85
HLA-DQA1	0.98	0.98	0.97
HLA-DQB1	0.97	0.98	0.99
HLA-DRB1	0.91	0.95	0.9
HLA-DRB3	0.94	0.96	0.96
HLA-DRB4	0.98	0.98	0.98
HLA-DRB5	0.99	1	1

Supplementary Table 3. Allele-specific sensitivity, specificity and positive predictive value (PPV) for alleles discussed in the main text. Statistics are calculated from 2/3 – 1/3 cross-validation experiment at a 4-digit level resolution, without applying a call threshold on posterior probabilities (call rate=100%), columns 2-4) , and with applying a threshold of T=0.7 (columns 5-7).

Allele	No call threshold			T=0.7		
	Sensitivity	Specificity	PPV	Sensitivity	Specificity	PPV
*A*02:01*	1	1	0.96	1	1	0.96
*B*44:02*	0.99	1	0.98	0.99	1	0.99
*B*38:01*	1	1	0.89	1	1	0.92
*B*55:01*	1	1	1	1	1	1
*DRB1*15:01*	1	1	0.98	1	1	0.98
*DRB1*13:03*	0.93	1	1	1	1	1
*DRB1*08:01*	1	1	0.84	1	1	0.84
*DRB1*03:01*	0.99	1	0.99	0.99	1	1
*DQA1*01:01*	1	1	1	1	1	1
*DQB1*03:02*	0.97	0.99	0.95	0.97	0.99	0.95
*DQB1*06:02*	0.99	1	0.98	0.99	1	0.99
*DRB1*01:01*	0.82	0.99	0.9	0.99	0.99	0.9

HLA allele/SNP	Parameters in full model	Effect	Freq in UK controls	FEM p-value	FEM OR	FEM 95% CIs	Q	P_heterogeneity	I^2	HLA alleles in LD (UK)	SNPs in LD (UK)
*DRB1*15:01*	2	additive	0.143	<1e-600	3.92	[3.74,4.12]	21.6	0.017	53.7	DQB1*06:02, DRB5*01, DRB5*09 (r^2>0.9), DQA1*01:02 (r^2>0.8)	rs3135391, rs3135388, rs3129889, rs9271366 (r^2>0.95)
		correction term for homozygotes		8.5E-22	0.54	[0.47,0.61]	14.6	0.149	31.3		
*A*02:01*	2	additive	0.276	7.8E-70	0.67	[0.64,0.70]	23.9	0.008	58.1	None	rs4713274, rs9295825 (r^2>0.9)
		correction term for homozygotes		3.3E-05	1.26	[1.13,1.41]	9.4	0.493	0		
*DRB1*13:03*	1	additive	0.009	6.2E-55	2.62	[2.32,2.96]	12.4	0.258	19.5	None	None
*DRB1*03:01*	2	additive	0.143	3.5E-08	1.16	[1.10,1.22]	15.0	0.132	33.4	DQA1*05:01, DQB1*02:01 (r^2>0.95)	rs1059615, rs2187668, rs2284189, rs9273327, rs2854275, rs3129716, rs2856674 (r^2>0.95)
		correction term for homozygotes		1.3E-30	2.58	[2.19,3.03]	19.1	0.039	47.6		
rs9277565_T	1	additive	0.206	2.1E-52	1.32	[1.27,1.36]	8.2	0.613	0	DPB1*03:01 (r^2>0.5)	rs9277561 rs9277567 (r^2>0.95)
*B*44:02*	1	additive	0.114	4.7E-17	0.78	[0.74,0.83]	9.0	0.528	0	C*05:01 (r^2>0.5)	rs9266773 (r^2>0.9)
*B*38:01*	1	additive	0.011	8.0E-23	0.48	[0.42,0.56]	8.6	0.571	0	None	None
*DRB1*08:01*	1	additive	0.02	1.0E-23	1.55	[1.42,1.69]	8.8	0.547	0	DQA1*04:01 (r^2>0.95), DQB1*04:02 (r^2>0.9)	rs7775055, rs4713586 (r^2>0.9)
*DQA1*01:01* (Interaction with *DRB1*15:01*)	1	additive effect of *DQA1*01:01* in the presence of *DRB1*15:01*	0.144	1.3E-17	0.65	[0.59,0.72]	19.4	0.036	48.3	DQB1*05:01, DRB1*01:01 (r^2>0.5)	rs13193645 (r^2>0.95)
rs2229092_C	1	additive	0.060	1.7E-22	1.33	[1.26,1.41]	23.3	0.01	57.1	None	None
B*55:01	1	additive	0.018	6.9E-11	0.63	[0.55,0.73]	10.2	0.42	2.2	None	rs3819284, rs3093547, rs9765960 (r^2>0.5)
DQB1*03:02	1	dominant	0.105	1.8E-22	1.30	[1.23,1.37]	21.4	0.019	53.2	DQA1*03:01 (r^2>0.5)	rs3957146, rs3998159, rs7454108, rs9275334, rs9275495, rs9275530, rs9275532 (r^2>0.95)
DQB1*03:01	1	Allelic interaction with DQB1*03:02	0.187	7.1E-12	0.60	[0.52,0.69]	7.0	0.723	0	None	rs5000632, rs9357152, rs9378125 (r^2>0.5)

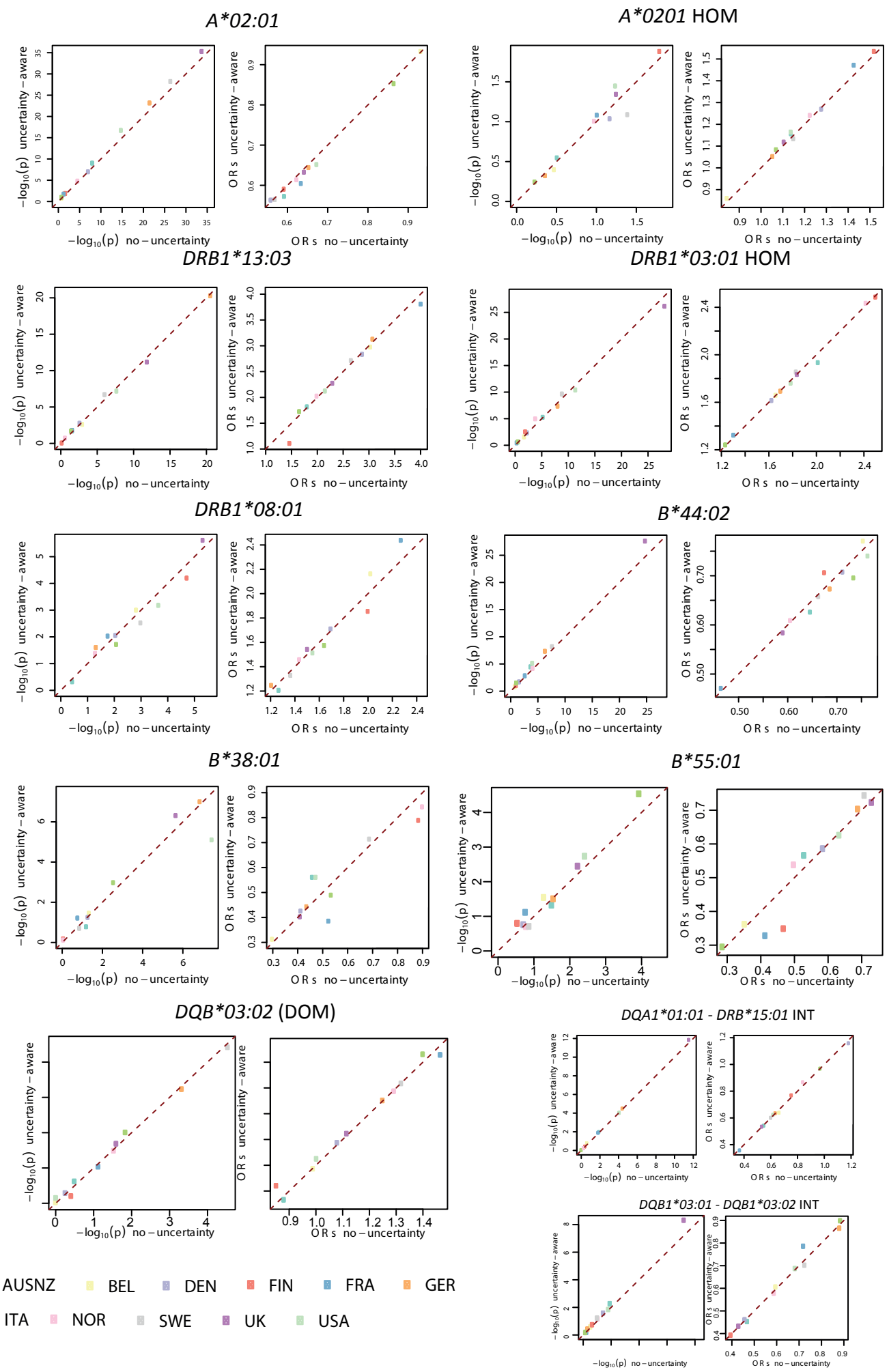
Supplementary Table 4. HLA alleles and SNPs found to be significantly associated with MS in this study. All alleles/SNPs are presented in the order of inclusion to the model (See Supplementary Material for details), with last three entries included after additional evidence from the mega-analysis. Second column contains the number of parameters used for each allele in the full model, while the third describes the effect assumed for each of the parameters (all effects are assumed in the log-odds scale). Columns 4-6 contain the P values, estimated odds ratios and 95% confidence intervals (CIs), respectively, of the alleles in the full model. Estimated values are from the fixed effect meta-analysis across cohorts. Columns 7,8 contain Cochrane's Q statistic for heterogeneity and the associated p-value. Column 9 contains I^2, an estimate for the proportion of total variability explained by heterogeneity. Last two columns contain the HLA alleles and SNPs, respectively, in LD with the alleles included in the full model (column 2). Entry "None" in these columns means that there is no allele with an r^2 of 0.5 or more with the reported allele. For HLA alleles (column 10), all alleles with an r^2>0.5 are reported. For SNPs (column 11), only the most associated ones are (with r^2 thresholds of 0.95, 0.9, 0.8 and 0.5).

Supplementary Table 5: Patient counts across cohorts for the secondary phenotypes analysed in this study.

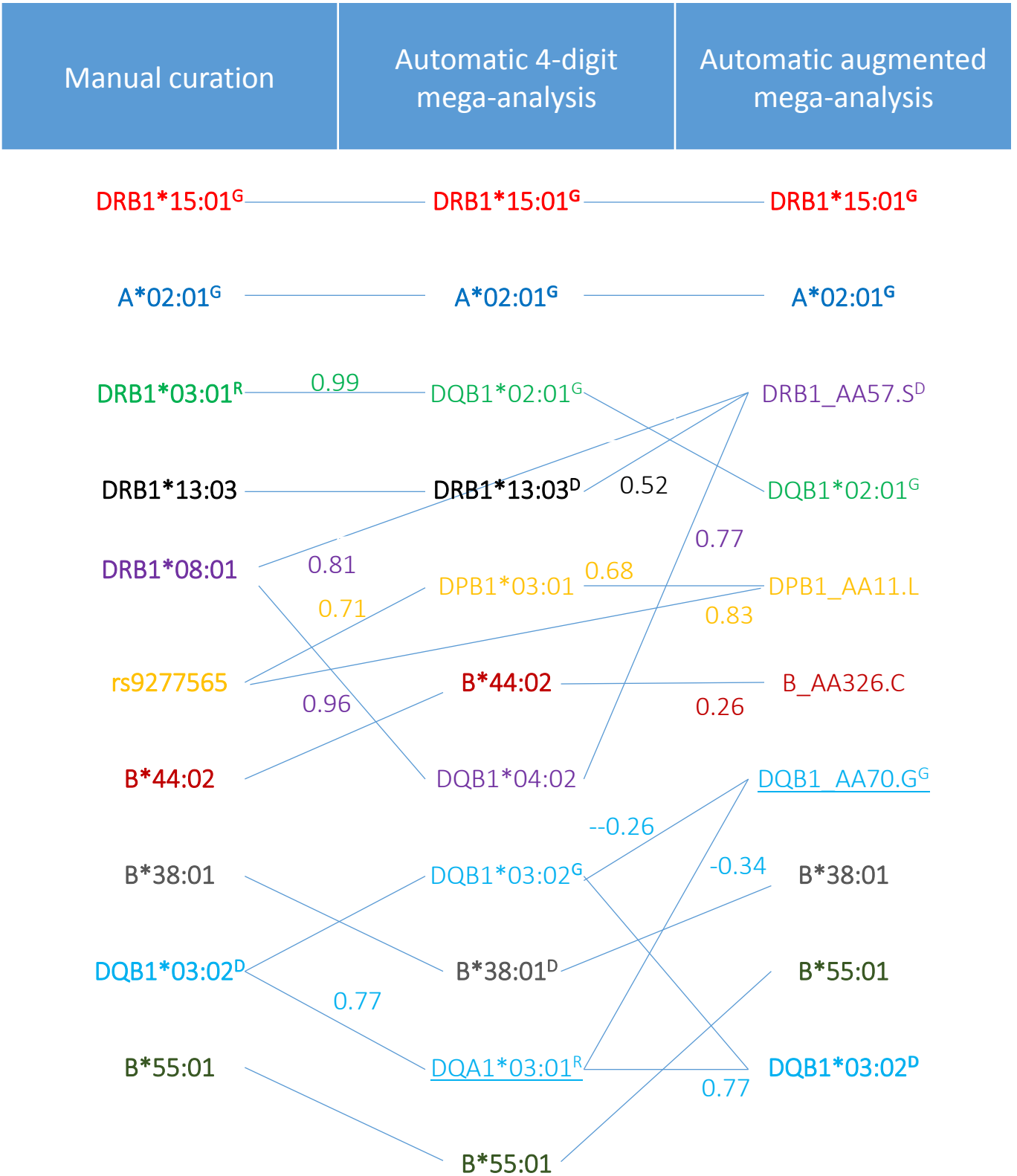
cohort	AAO	Clinical Course	MSSS	MSSS<2.5	MSSS>7.5
AUSNZ	794	810	713	177	151
BEL	298	295	271	79	80
DEN	889	890	885	231	138
FIN	275	290	246	63	64
FRA	366	366	360	93	69
GER	391	2046	352	130	30
ITA	910	924	833	439	52
NOR	812	880	406	123	111
SWE	2375	2298	1998	699	381
UK	3909	3651	2001	436	531
USA	1122	1141	1043	407	161

Supplementary Table 6: Comparison of models for multiple sclerosis disease risk and age at onset (AAO). For disease risk, all other effects reported in Supplementary Table 4 are included. All models have the same number of parameters, hence the results are equivalent for AIC and BIC. The model with the highest likelihood is identified in orange and the model with the second highest likelihood is identified in yellow.

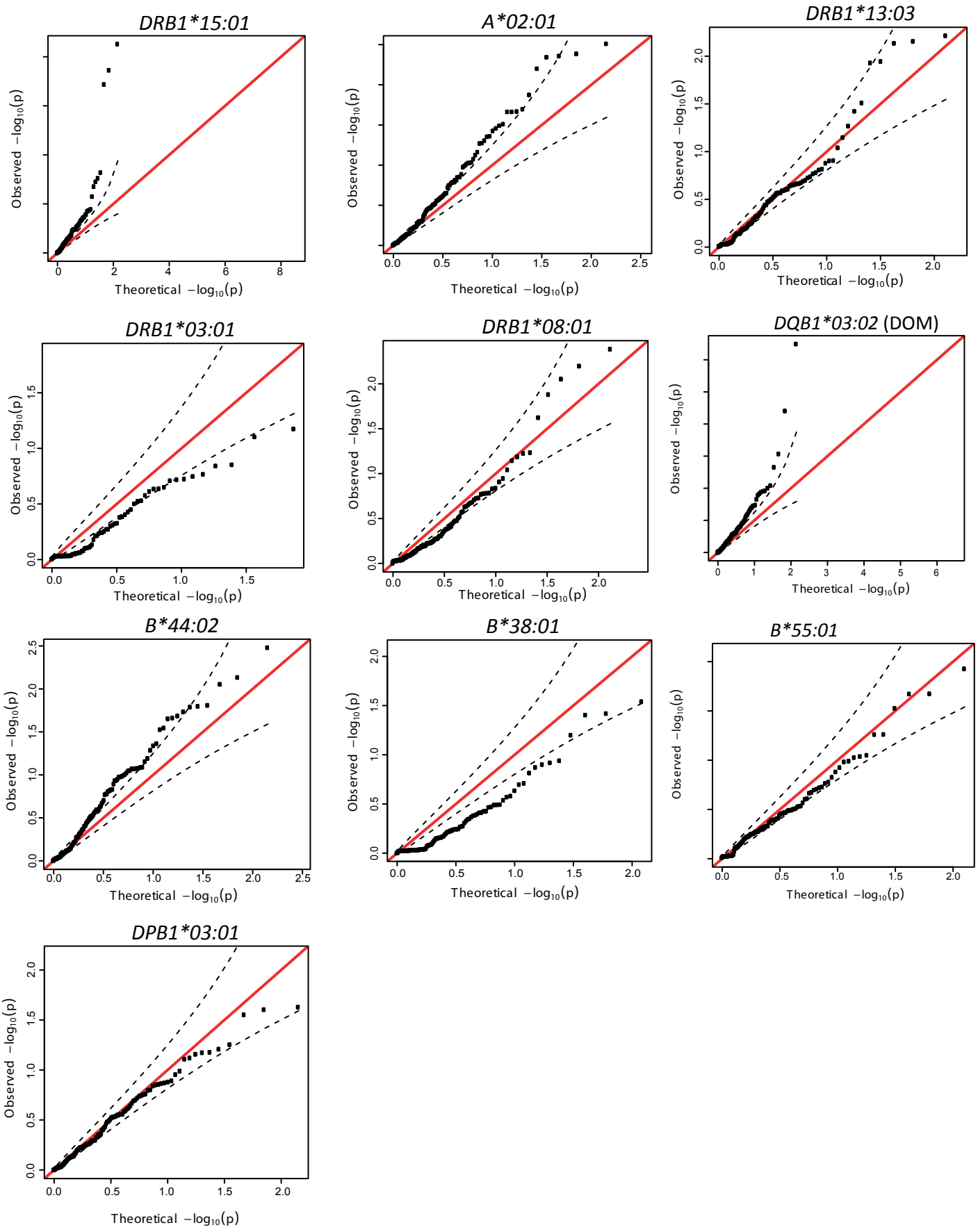
Main effect	Second effect	Second locus effect	Disease Risk		Age at Onset	
			AIC in UK	AIC in All	AIC in UK	AIC in All
*DRB1*15:01*	*DQA1*01:01*	Additive	15751	56435	29296	90472
*DRB1*15:01*	*DRB1*01:01*	Additive	15757	56437	29294	90467
*DRB1*15:01*	DQB1_AA70.G	Hom	15733	56382	30130	85297
*DRB1*15:01*	*DQA1*01:01*	Interaction	15722	56378	30135	85316
*DRB1*15:01*	*DRB1*01:01*	Interaction	15734	56413	30134	85315
*DRB1*15:01*	DQB1_AA70.G	Hom interaction	15727	56354	30131	85305
*DQB1*06:02*	*DQA1*01:01*	Additive	15829	56730	29296	90471
*DQB1*06:02*	*DRB1*01:01*	Additive	15837	56737	29293	90466
*DQB1*06:02*	DQB1_AA70.G	Hom	15825	56701	30130	85296
*DQB1*06:02*	*DQA1*01:01*	Interaction	15812	56695	30135	85316
*DQB1*06:02*	*DRB1*01:01*	Interaction	15821	56723	30134	85314
*DQB1*06:02*	DQB1_AA70.G	Hom interaction	15826	56696	30131	85305



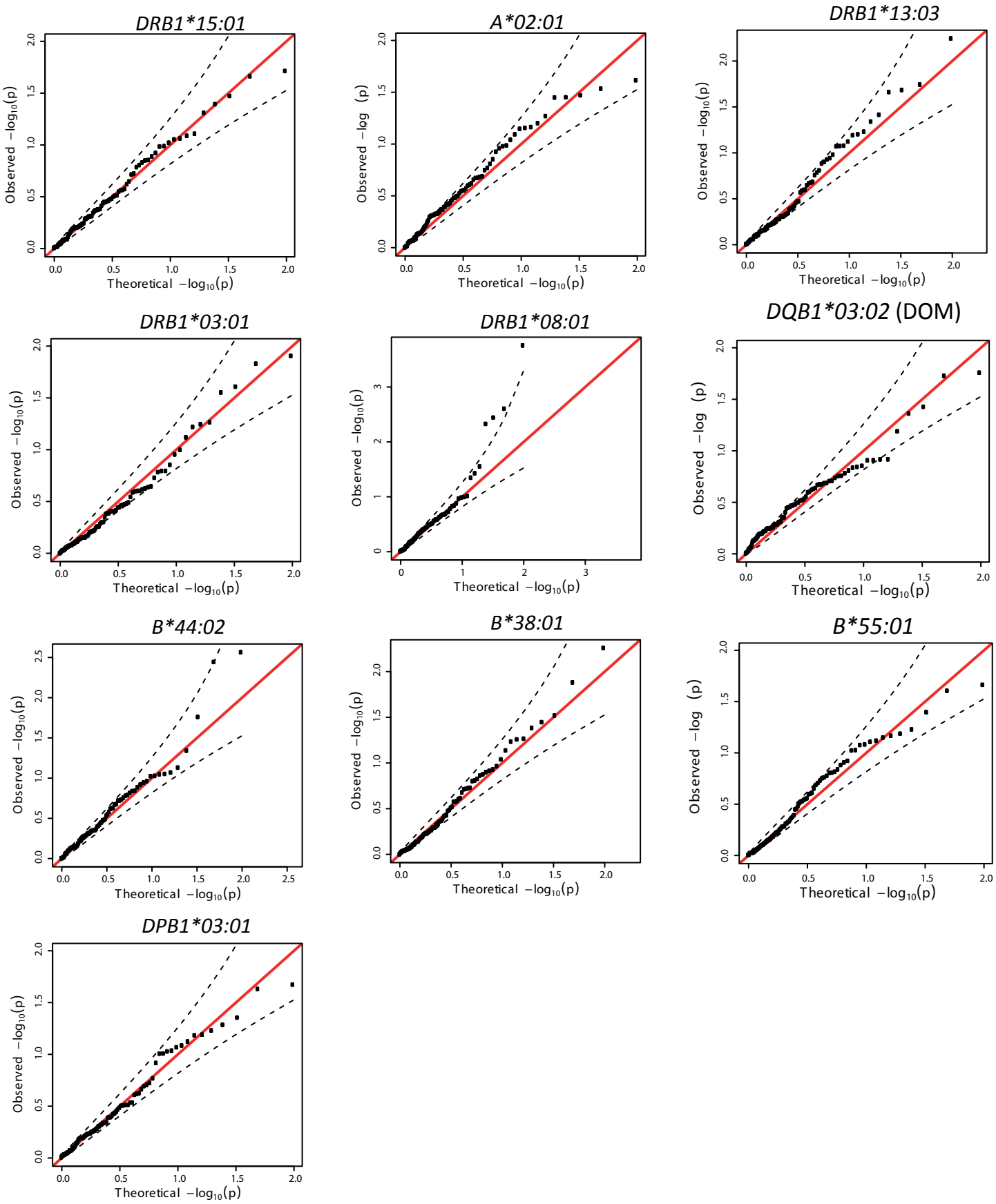
Supplementary Figure 1. Effects of incorporating uncertainty in classical HLA allele prediction. Comparison of P values and odds ratios estimated for each cohort for all classical HLA allele effects shown in Fig. 1 under methods taking account of uncertainty in classical HLA allele imputation (y-axis) and not taking into account of uncertainty (x-axis). Values reported are from a model including additive and non-additive effects for *DRB1*15:01* and PCs, in addition to the indicated allele.



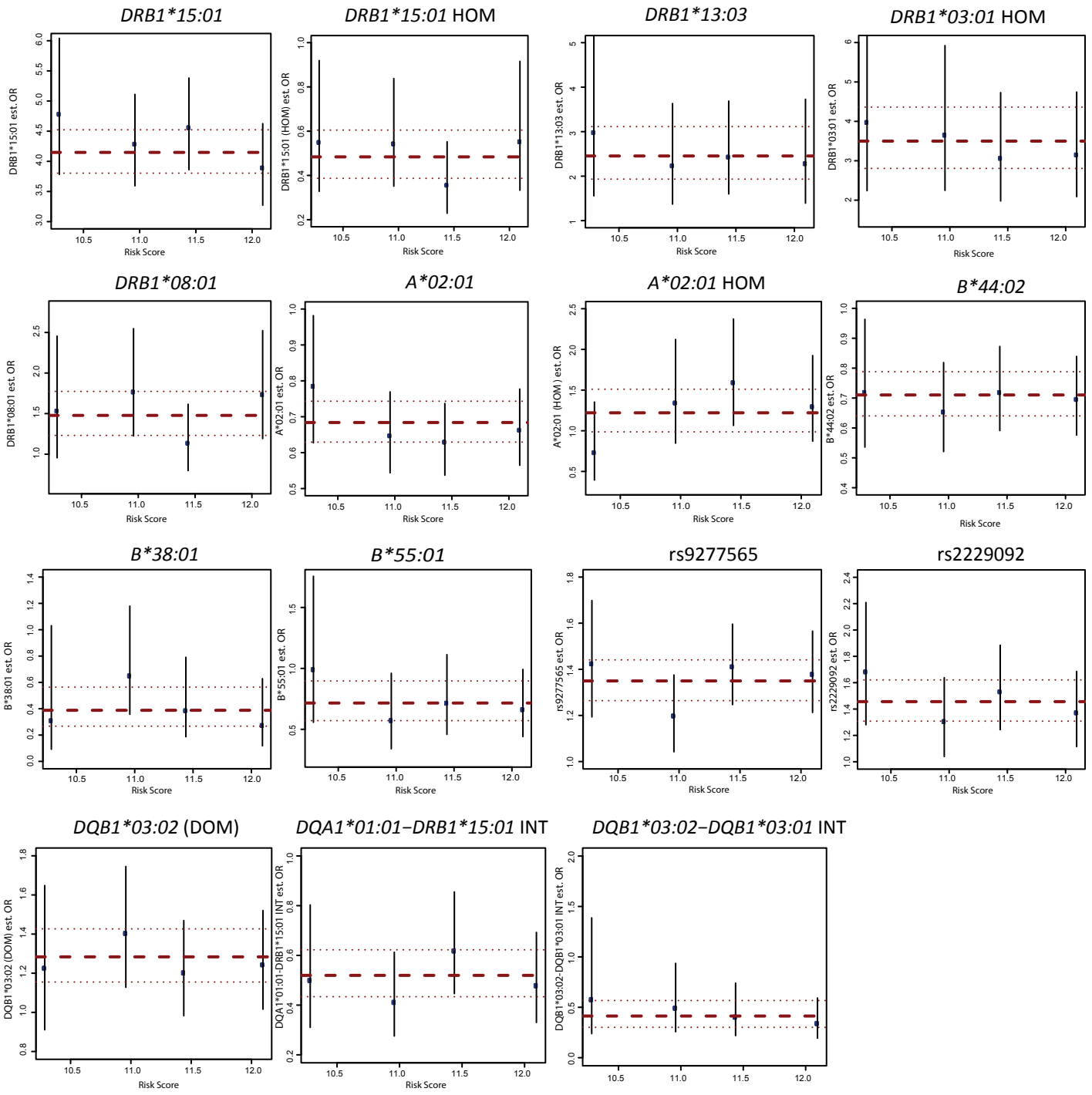
Supplementary Figure 2. Summary of the results from the model selection procedures (without interaction search) using manual curation (left), automatic model search with alleles at 4-digit resolution (middle) and automatic mega-analysis augmented with alleles grouped at 2-digit resolution and by sharing of amino acid residues. Alleles shown in order of inclusion in each approach and matched between approaches by linkage disequilibrium (Pearson correlation coefficient across entire cohort shown and groups identified by colours). At each stage, models with additive (no superscript), dominant (D), recessive (R) and general (G) risk were tested. The allele used to refer to each group is indicated by bold. Effects identified that are better explained by interactions between other alleles are underlined. Note that DRB1_AA57.S is a near-perfect proxy (r = 0.95) for the presence of either *DRB1*08:01* or *DRB1*13:03*, though a model with independent effects for the two alleles provides better fit. We note also that B_AA326.C is weakly correlated with the presence of *B*44:02* (r = 0.26), though the latter remains significant in FEM if the former is included.



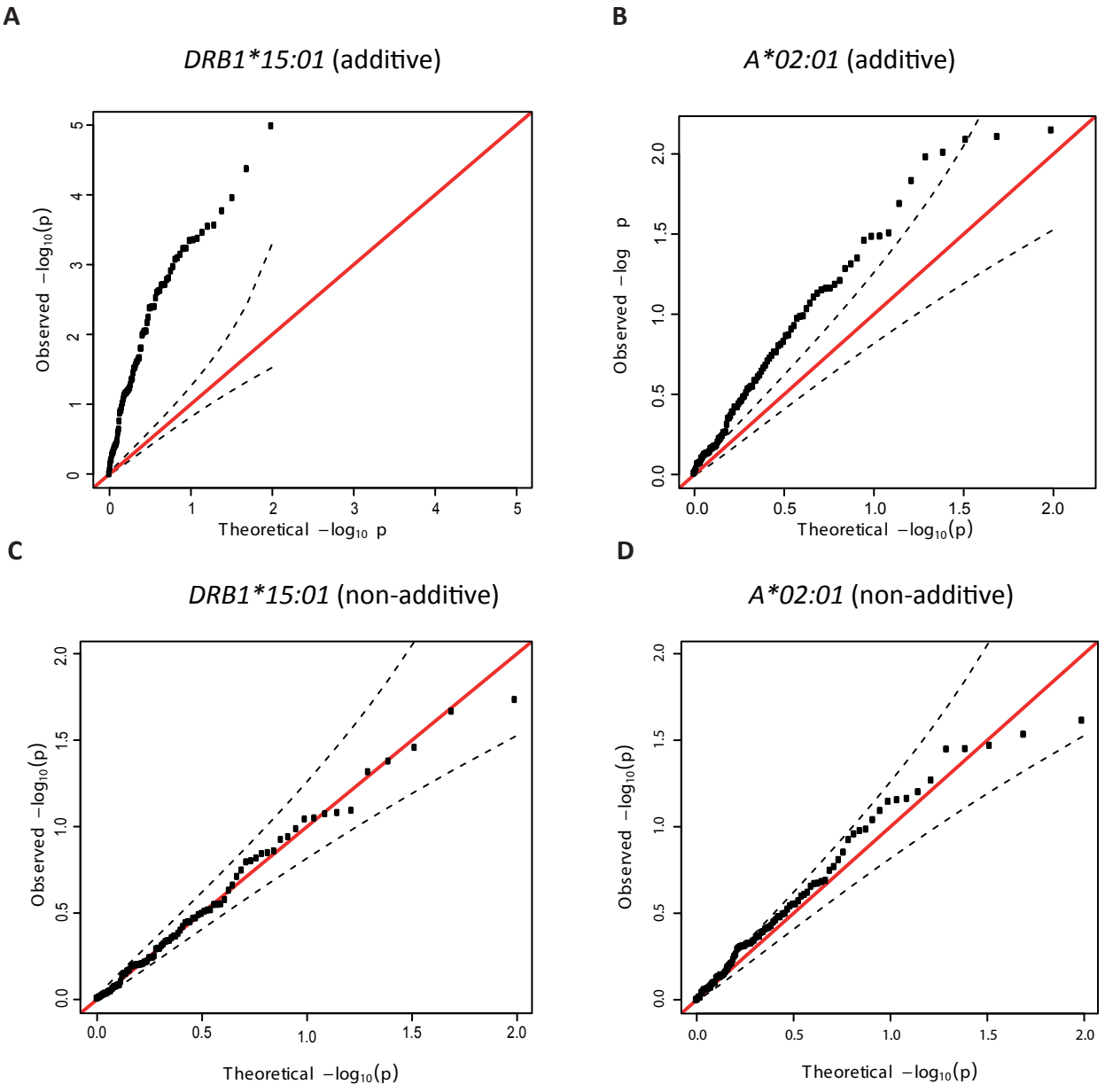
Supplementary Figure 2. Evidence for interactions among classical HLA alleles affecting risk for multiple sclerosis. QQ plots showing the distribution of P values for the interaction term between HLA alleles reported in the full model and other classical HLA alleles. For each test, a model including non-additive effects for *DRB*15:01* and *A*02:01* is fitted, along with all other risk HLA alleles and SNPs, and with the potential interaction (see section 7 of the supplement for model details).



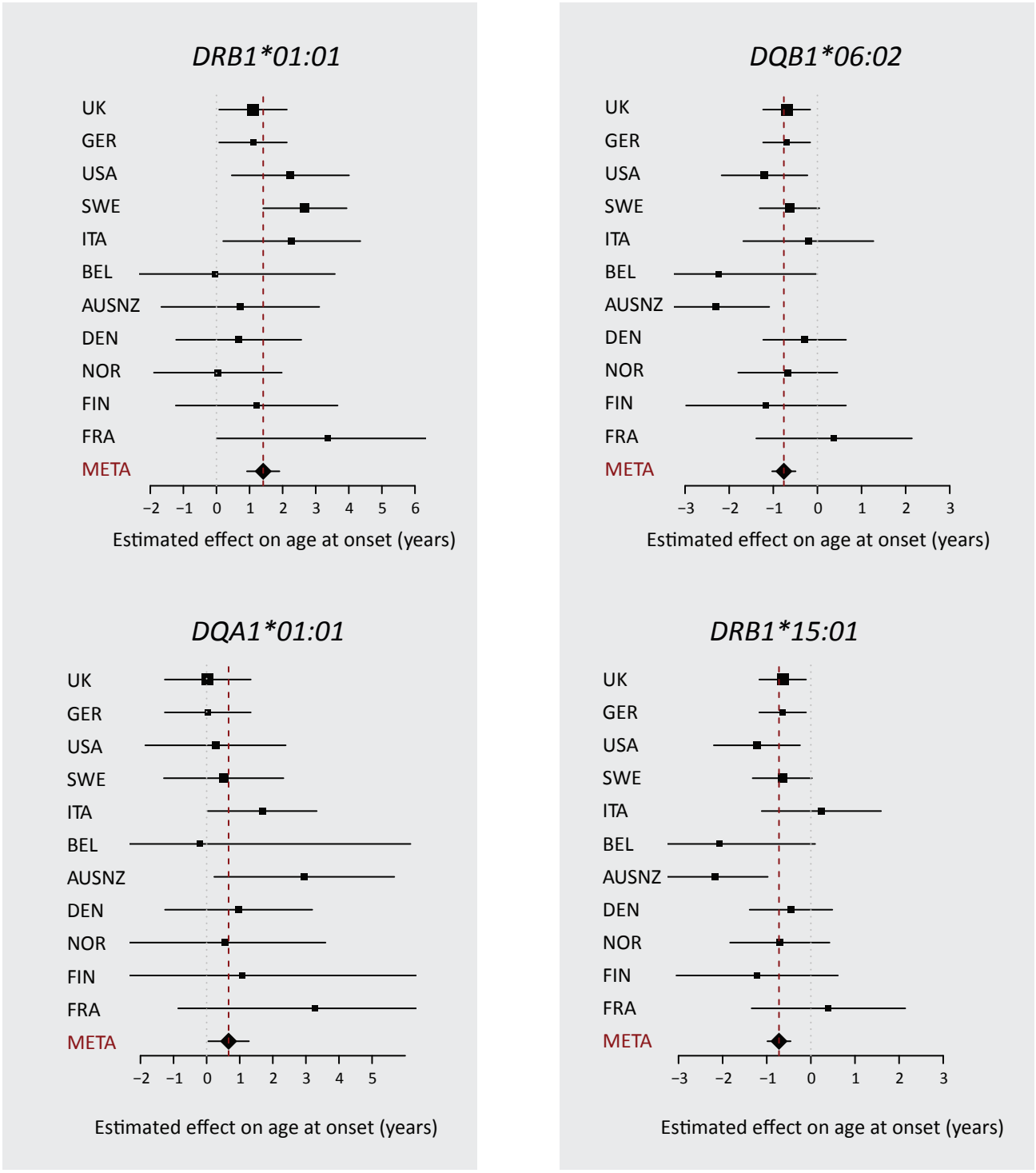
Supplementary Figure 4. Evidence for interactions between classical HLA alleles and non-HLA disease-associated variants affecting risk for multiple sclerosis. QQ plots showing the distribution of P values for the interaction term between HLA alleles reported in the full model and non-HLA SNP variation influencing genetic risk for multiple sclerosis (from Ref. S1). For each associated HLA allele, a model including non-additive effects for *DRB*15:01* is fitted, along with the potential interaction (see section 8 of the supplement for model details).



Supplementary Figure 5. Interactions between HLA variants and combined non-HLA risk for multiple sclerosis. The effect of the indicated allele or variant among individuals stratified in to quartiles by a combined non-HLA risk score, obtained by multiplying odds-ratios associated with each genotype an individual carries at non-HLA loci influencing risk (from Ref. S1). The point estimate and 95% confidence intervals for effect size are estimated independently for each quartile of non-HLA genetic risk. Dashed and dotted lines indicate the combined point estimate and 95% confidence interval respectively. All analyses shown here are for the UK cohort only.



Supplementary Figure 6. Conflation between evidence for interactions and a departure from additivity for *DRB1*15:01* and *A*02:01*. (A, B) QQ-plots showing the distribution of P values for interaction terms between DRB1*15:01 (A) and *A*02:01* (B), and non-HLA variants known to affect risk of multiple sclerosis (from Ref. S1) under a model where the HLA allele acts additively (on the log-odds scale). **(C,D)** As for A,B, but where a term capturing departures from additivity is included for *DRB1*15:01* and *A*02:01*. Because potential interactions are modelled as an additional additive effect of the associated SNP that only acts in the presence of at least one copy of the HLA allele, deviations from additivity for the HLA effects can partly mimic potential interactions with the non-HLA variants.



Supplementary Figure 7. Classical HLA alleles affecting age-at-onset.
Meta-analysis showing effects of pairs of classical HLA alleles in strong linkage disequilibrium (*DRB1*15:01 – DQB1*06:02* and *DRB1*01:01 – DQA1*01:01*) on age-at-onset.

Supplementary Note: Estimating genome-wide effect of epistasis on the HLA loci using linear mixed models

Abstract

To test the total contribution of epistasis to the effect sizes of HLA alleles, we develop a statistical framework that uses a linear mixed model (LMM) to model the total net impact of interaction between the HLA allele and an unmeasured polygenic score. We can then use realised relatedness matrices of "unrelated" (i.e. not closely related) individuals, estimated from genome-wide SNP data, to fit these models (analogously to how LMMs can be used to estimate the total contribution of common SNPs to additive heritability e.g. [6]). In this appendix we outline the basic model (in which effect sizes are treated as unobserved heritable traits), demonstrate that this model induces a heritability of HLA alleles within cases proportional to the strength of total epistasis, and give equations to convert between the size of this induced heritability and the parameters of the epistatic model. We show using simulations that this method is well calibrated, and has high power to detect moderate polygenic epistasis in studies with at least 3000 cases and 3000 controls. Finally, we apply this model to the IMSGC Immunochip data and imputed HLA alleles to calculate the total contribution of epistasis to the HLA odds ratios from variants included on the Immunochip.

A.1 The personalised effect size model

Assume that a continuous phenotype is influenced by a marginal effect of the dosage of a major locus x_i (such as the HLA), with a marginal effect size $\tilde{\beta}$, as well as additive/additive interactions between the major locus and L minor loci (e.g. non-HLA loci) with normalized dosages a_1 to a_L, each with no marginal effect and each increasing the effect size of the major locus by β_{xj}. Normalized in this sense means $a_j = \frac{a'_j - 2f_j}{\sqrt{2f_i(1-f_i)}}$, where $a'_j \in (0,1,2)$ is the unnormalized dosage and f_j is the allele frequency. We can write down the overall genetic score for individual i as

$$y_i = \beta_0 + \tilde{\beta} x_i + \sum_{j=1}^{L} \beta_{xj} x_i a_{ji} \tag{1}$$

We can rewrite this as

$$\begin{aligned} y_i &= \beta_0 + x_i \left(\tilde{\beta} + \sum_j \beta_{xj} a_{ji} \right) \\ &= \beta_0 + x_i \beta_i \end{aligned} \tag{2}$$

where $\beta_i = \tilde{\beta} + \sum_j \beta_{xj} a_{ji}$ is now a (potentially unobserved) personalised effect size, equivalent to the amount that the phenotype is increased by the major locus in individuals with a genome identical to that of individual i.

If we look across a number of individuals, providing that L is large and that the β_{xj} values are drawn from a distribution that is close to normal, the vector of personalised effect sizes will be distributed according to

$$\vec{\beta} \sim N(\tilde{\beta}, \Sigma v^2) \tag{3}$$

where $v^2 = \sum_j \beta_{xj}^2$ and Σ is the realised relatedness matrix of the individuals. This is defined as the average covariance in standardized genotypes a_{ij} (as defined by[6]) across the set of SNPs that interact with the major locus. As we do not know the true set of SNPs that are interacting with the major locus, we instead use a set of genome-wide SNPs to estimate this relatedness matrix.

A.2 Binary traits and within case analyses

For binary disease traits, we define a link function g between the continuous score and the probability of disease, such that

$$P(d_i = 1|y_i) = g(y_i) = g(\beta_0 + \beta_i x_i). \tag{4}$$

Usually $g(y_i) \to 1$ as $y_i \to \infty$, and $g(y_i) \to 0$ as $y_i \to -\infty$. For instance, if $g(y_i) = logit(y_i)$, then we have a logistic model with β_i corresponding to a personalised log-odds ratio. β_0 needs to be picked such that $E[g(y_i)] = K$, where K is the prevalence of the disease.

We can reverse this model in order to remove (usually unobserved) β_i, and instead create an (approximate) linear mixed model that predicts the dosage of the major allele using the case-control status and the realised relatedness matrix. Assuming that cases are rare in the population, or that controls are population controls, most of our information will come from within cases, and we thus create a within-case model:

$$\vec{x}|(\vec{d} = 1) \sim N\left[\mu_x, \sigma_x^2\left(h_x^2\Sigma + (1 - h_x^2)I\right)\right] \tag{5}$$

In this equation h_x^2 is the induced heritability of the major locus dosage in cases. As we show below, this in heritability is induced in cases by the genome-wide polygenic epistasis, and in the remainder of this section we describe how the relate this heritability parameter to the original model parameters.

The probability of a given case having a particular dosage x_i at the major locus, conditional on the case's personalised effect size, is

$$P(x_i|\beta_i, d_i = 1) = \frac{g(\beta_0 + \beta_i x_i)p(x_i)}{\sum_{x_i'} g(\beta_0 + \beta_i x_i')p(x_i')}. \tag{6}$$

The expected dosage of the risk allele in cases is then

$$\begin{aligned} \mu_x &= E[x_i|d_i = 1] \\ &= \int_{\beta_i}\left(\sum_{x_i} x_i P(x_i|\beta_i, d_i = 1)\right)P(\beta_i)d\beta_i, \end{aligned} \tag{7}$$

where $P(\beta_i)$ is the normal density with mean $\tilde{\beta}$ and variance v^2 (taken from equation A2).

The variance in risk allele dosage in cases is given by

$$\begin{aligned} \sigma_x^2 &= Var[x_i|d_i = 1] \\ &= \int_{\beta_i}\left(\sum_{x_i} x_i^2 P(x_i|\beta_i, d_i = 1)\right)P(\beta_i)d\beta_i - E[x_i|d_i = 1]^2 \end{aligned} \tag{8}$$

The covariance in major locus dosage between two individuals i and j with a kinship coefficient Σ_{ij} can be calculated using

$$\begin{aligned} E[x_i x_j|d_i = 1, d_j = 1, \Sigma_{ij}] &= \int_{\beta_i}\int_{\beta_j}\left(\sum_{x_i}\sum_{x_j} x_i x_j P(x_i = 1|\beta_i, d_i = 1)P(x_j|\beta_j, d_j = 1)\right)P(\beta_i, \beta_j|\Sigma_{ij})d\beta_i d\beta_j \\ Cov[x_i, x_j|d_i = 1, d_j = 1] &= E[x_i x_j|d_i = 1, d_j = 1, \Sigma_{ij}] - E[x_i|d_i = 1]^2 \end{aligned} \tag{9}$$

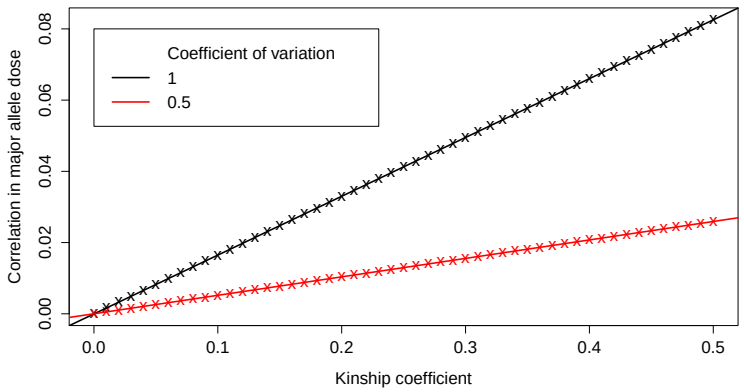


Figure A1: Relationship between kinship coefficient and correlation in major allele count in cases for different coefficients of variation. $K = 0.01$, $f = 0.3$, $\beta = log(2)$. Crosses represent the actual values, and the lines show a linear fit by least squares.

where $P(\beta_i, \beta_j|\Sigma_{ij})$ is a multivariate normal density, with mean, variance and covariance taken from equation 3.

The heritability h_x^2 can be estimated by calculating the covariance for various kinship values, and fitting a linear model where

$$\frac{Cov[x_i, x_j|d_i = 1, d_j = 1]}{Var[x_i|d_i = 1]} = h_x^2 \Sigma_{ij} \tag{10}$$

For some choices of link function the true covariance will not be linearly dependent on the kinship coefficient, and so this linear will be an approximation. To test how close this approximation is, we calculated model parameters under the logit link model, assuming a major allele with a risk allele frequency of 30% and an odds ratio of 2 for a disease with a prevalence of $K = 0.01$. We varied the disease prevalence and the size of the epistatic effect (measured by the coefficient of variation $\frac{v}{\beta}$).

For these parameters the logit model induced a additive heritability of major allele in cases, in the sense that the correlation in major allele count was very close to linearly proportion to the coefficient of relatedness (Figure A1).

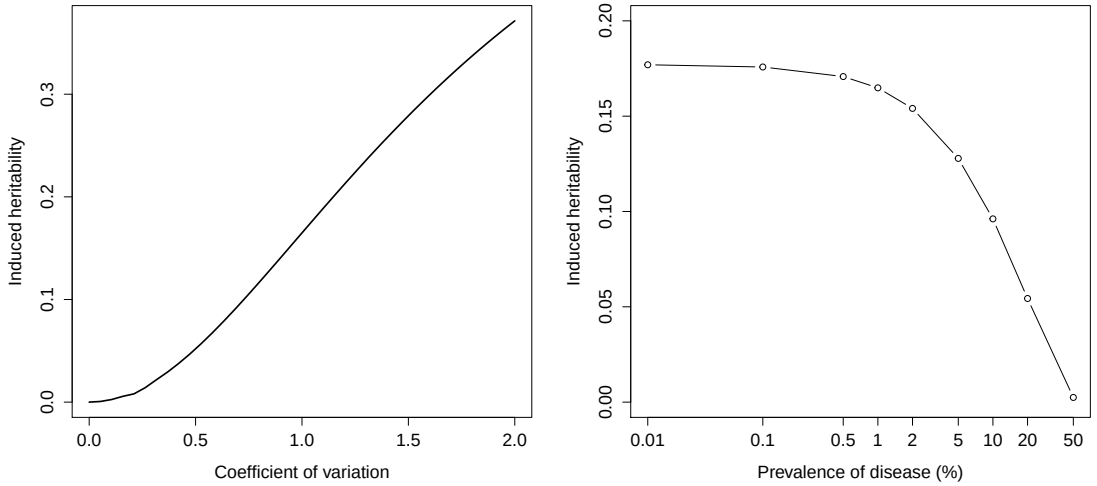
A.3 Strength of the induced heritability

To investigate the strength of the induced heritability, we calculated model parameters under the logit link model. As above, we looked at a major allele with a risk allele frequency of 30% and an odds ratio of 2, and varied the coefficient of variation $\frac{v}{\beta}$.

For a relatively rare disease ($K = 0.01$) genome-wide epistasis creates a detectable heritability of major alleles in cases ($h_x^2 > 0.1$) when the coefficient of variation rises around 70% or so (Figure A2a). For common diseases ($K > 0.02$) the induced heritability falls off rapidly (Figure A2b).

A.4 Correcting for population stratification

HLA alleles are highly stratified geographically, and this population stratification could also contribute to a heritability of HLA alleles (as geographically close individuals will have both more similar HLA alleles and higher genome-wide relatedness). However, if cases and controls are well matched, this heritability due to stratification will be similar in cases and in controls. Thus to correct for stratification we calculate the heritability of HLA alleles in both cases and controls, and subtract the with-control heritability from the with-case heritability to estimate the heritability induced by epistasis h_x^2.



(a) $K = 0.01$, $f = 0.3$, $\beta = log(2)$

(b) $f = 0.3$, $\tilde{\beta} = log(2)$, $\frac{v}{\tilde{\beta}} = 1$

Figure A2: Induced heritability of major allele count in cases under a logit link for different coefficients of variation and prevalence.

A.5 Comparison with GCTA's GxqE method

Our method above is similar to the gene-environment interaction methods included in GCTA [7], and in particular with the GxqE method for estimating a variance component due to an interaction between polygenic risk and a quantitative environmental exposure. This method estimates variance components by fitting the following model:

$$\vec{d} \sim N\left(\beta_0 + \tilde{\beta}\vec{x}, V_G\Sigma + V_{GxqE}\Sigma' + V_e I\right). \tag{11}$$

where $\Sigma'_{ij} = \Sigma_{ij}x_i x_j$ (i.e. the product of the genome-wide covariance and covariance in environmental exposures). As with standard variance component methods for discrete diseases [2], the binary disease status is approximated as a normally distributed variable.

To see the connection between this random effects model and our own, let us extend our model given in equation 2 in a similar fashion. We will model the disease status as normal variable given by a combination of the major locus effect y_i (including both marginal and epistasis effects, as in equation 2), an additive polygenic component $l_i \sim N(0, V_G\Sigma)$ and an error term $e_i \sim N(0, V_e I)$, such that

$$\begin{aligned} d_i &= y_i + l_i + e_i \\ &= \beta_0 + \beta_i x_i + l_i + e_i. \end{aligned} \tag{12}$$

Assuming no covariance between l_i and β_i (i.e. that effect sizes for additive genetic risk loci and gene-environment interaction loci are independent) the covariance in disease status is then given by

$$\begin{aligned} Cov[d_i, d_j] &= x_i x_j Cov[\beta_i, \beta_j] + Cov[l_i, l_j] + Cov[e_i, e_j] \\ &= x_i x_j v^2 \Sigma_{ij} + V_G\Sigma_{ij} + \delta_{ij}V_e \end{aligned} \tag{13}$$

where δ_{ij} is the delta function ($\delta_{ij} = 1$ if $i = j$ and 0 otherwise).

By comparison with equation 11, we can see that $V_{GxqE} = v^2$. This means that if the environmental exposure is substituted for the dosage of a major risk locus, the GxqE variance component

estimated by GCTA is also an estimate of the epistatic variance due to polygenic interaction with that major locus.

In essence, both GCTA's GxqE method, and our induced heritability method, aim to infer the same parameter. However, while our method uses an explicit link function to model case-control status as a binary variable, and applies the linear mixed model approximation to the conditional allele dosage $x_i|d_i$, the GxqE method applies the approximation directly to the disease status d_i.

A.6 Simulation study

To validate that the method works as expected, and to investigate the power of the method, we carried out a simulation study. We calculated the power and distribution of estimated induced heritabilities and p-values across simulations for different values of the simple size N and epistatic variance v^2. In all cases we simulated a disease with $K = 0.01$, an HLA locus with allele frequency $f_{HLA} = 0.143$ and effect size $\beta_{HLA} = log(3.78)$ (i.e. the observed allele frequency and effect size for HLA-DRB1*15:01). We assumed that Immunochip variants accounted for 20% of the additive variance of the disease on the liability scale.

We generated a large simulated Immunochip test population of 1 million individuals to test the method (i.e. a large enough population to see significant numbers of affected individuals even with $K = 0.01$). We constructed this set using seed data from 8783 QC+ UK controls with 161119 QC+ SNPs. We first phased each chromosome using SHAPEIT2[3] (r790) with default parameters, and we then generated new simulated individuals from these haplotypes using HAPGEN[5] (v2.2.0), again with default settings.

For a given simulation, we picked 1000 simulated causal variants from the set of all non-HLA Immunochip variants with a minor allele frequency greater than 0.1%. For each locus g we assigned an additive effect size β_g^{add} and an epistatic effect size β_g^{epi}. β_g^{add} were drawn from $N(0, \sigma_{add}^2)$, where $\sigma_{add} = 0.046$ was chosen such that these variants explained 20% of the additive variance on the liability scale. β_g^{epi} were drawn from $N(0, v^2)$, as we drawn independently of β_g^{add}. Finally, we assigned an HLA dosage to each individual, drawn from $binomial(2, f_{HLA})$.

We then assigned each individual as a case with a probability equal to

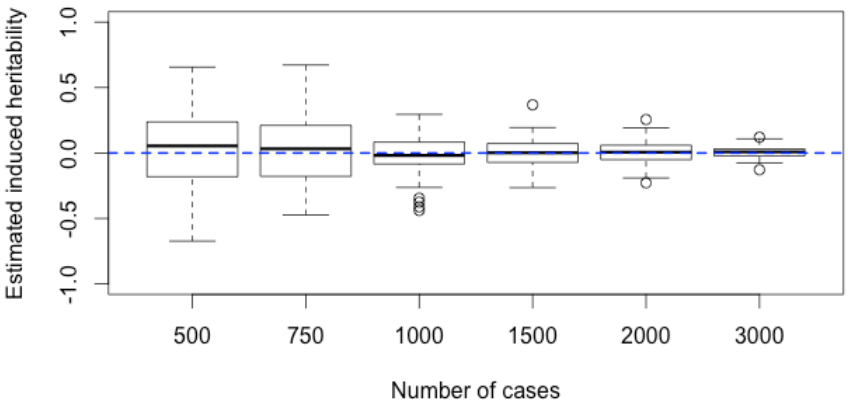
$$P = logit\left(\mu + x_{HLA}(\beta_{HLA} + \sum_g \sqrt{2f_g(1-f_g)}\beta_g^{epi}x_g) + \sum_g \sqrt{2f_g(1-f_g)}\beta_g^{add}x_g)\right) \qquad (14)$$

where x_{HLA} and x_g are genotype dosages at HLA and non-HLA loci respectively, and where μ was selected such that $E[P(D)] = K$. To produce the final simulated dataset we randomly selected N of these cases and N controls (i.e. non-cases), and then calculated the induced heritability of the HLA allele in cases and controls. We repeated the whole process 100 times (starting with drawing the 1000 causal variants) for each tested value of N and v^2.

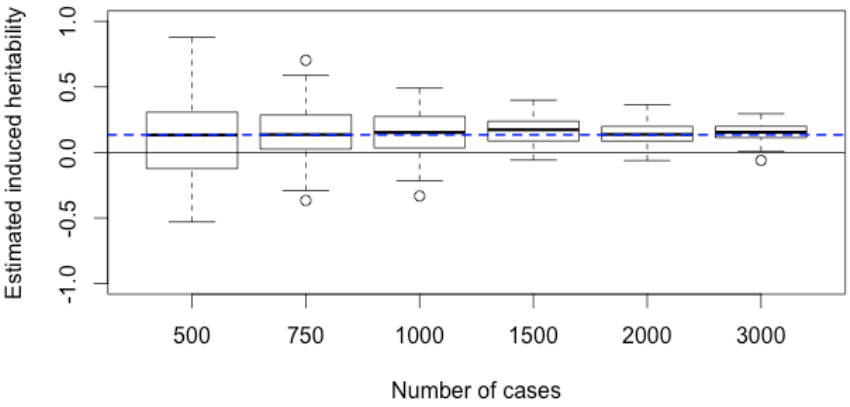
The estimated induced heritabilities across the simulations are shown in Figure A3. The estimated parameter is distributed around zero in the purely additive case (i.e. when $v = 0$, Figure A3a), and the distribution of p-values is uniform. In the presence of epistasis (Figures A3b and A3c) the estimates of the induced heritability are centered on the true heritability expected from the equations given in section A.2 above, and the range of the estimates tighten around the true value as the sample size increases. The power calculations (Figure A4) show that for strong epistasis ($v/\beta_{HLA} = 1$) you only need around 1500 cases and 1500 controls to have over 50% power to detect epistasis at a significance threshold of $p < 0.01$. To have a high power to detect moderate epistasis ($v/\beta_{HLA} = 0.5$), you need a study of at least 3000 cases and 3000 controls.

A.7 Application to Immunochip data

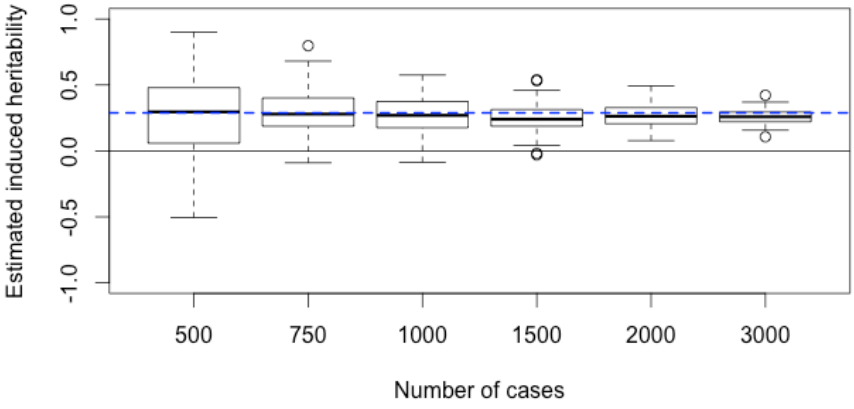
We calculated separate realised relatedness matrices for all cases and all controls within each cohort, using Plink 1.9. We removed all individuals with a high degree of relatedness ($\hat{\pi} > 0.2$), and used all SNPs that passed QC (described in supplementary methods section 1), excluding



(a) $v/\beta_{HLA} = 0$



(b) $v/\beta_{HLA} = 0.5$



(c) $v/\beta_{HLA} = 1$

Figure A3: The estimated induced heritability for three different coefficients of variation. The blue line shows the expected induced heritability using equations 9 and 10.

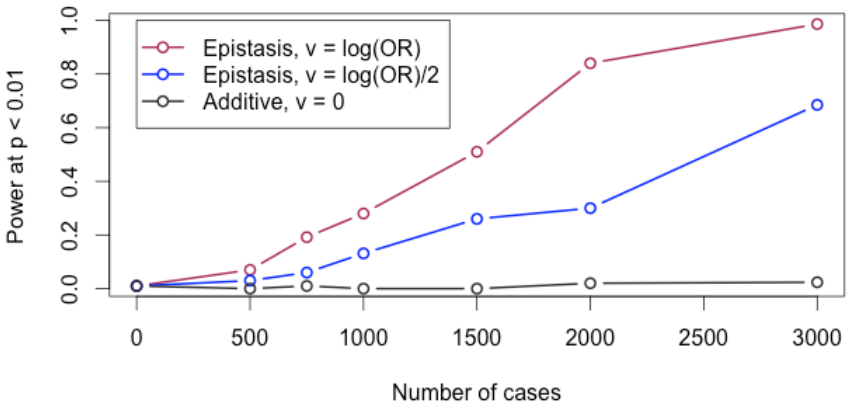


Figure A4: Power curves for detecting different degrees of epistasis on an HLA allele from simulations.

chromosome 6 (to avoid confounding by variants in LD with HLA alleles). We fitted the linear mixed models using GCTA v1.21 [7], with no constraints on the parameter estimates. For each combination of cohort, HLA allele and case-control status we estimated the heritability of the dosage of the HLA allele. The stratification-corrected induced heritabilities for each HLA allele across all cohorts were then meta-analysed under a fixed-effect model using the R package "meta" [4], and final results less than zero were truncated at zero. These results were converted from induced heritabilities to coefficients of epistatic variation using the equations in Section A.2 above. The results are shown in Figure 4 of the main paper. For all alleles the 95% confidence interval for the coefficient of variation overlapped zero.

For comparison, we also generated the heritability explained by the epistatic variance component $\frac{V_{GxgE}}{V_P}$ given by GCTA (again, calculated within cohorts, meta-analysed using "meta" and truncated at zero). The results are given below; again, the confidence intervals for all alleles overlapped zero.

Allele	Estimate	95% CI
A*02:01	0	0 - 0.008
B*38:01	0	0 - 0.113
B*44:02	0	0 - 0.017
DRB1*03:01	0.015	0 - 0.031
DRB1*08:01	0.046	0 - 0.090
DRB1*13:03	0.056	0 - 0.123
DRB1*15:01	0.041	0 - 0.123

Note that the parameters calculated do not actually represent the total genome-wide contribution to epistasis, as they only include SNPs present on the Immunochip. However, the Immunochip captures a large proportion of genome-wide genetic variation for diseases of immunity. For instance, Yang *et al* [1] found that >70% of genetic variation in inflammatory bowel disease risk tagged by GWAS arrays was also tagged by the Immunochip. Thus, while the contribution of epistasis is likely underestimated, the degree of underestimation should be small.

References

[1] G. B. Chen, S. H. Lee, M. J. Brion, G. W. Montgomery, N. R. Wray, et al. Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. *Hum. Mol. Genet.*, 23(17):4710–4720, Sep 2014.

[2] S. H. Lee, N. R. Wray, M. E. Goddard, and P. M. Visscher. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.*, 88(3):294–305, Mar 2011.

[3] J. O'Connell, D. Gurdasani, O. Delaneau, N. Pirastu, S. Ulivi, M. Cocca, M. Traglia, J. Huang, J. E. Huffman, I. Rudan, R. McQuillan, R. M. Fraser, H. Campbell, O. Polasek, G. Asiki, K. Ekoru, C. Hayward, A. F. Wright, V. Vitart, P. Navarro, J. F. Zagury, J. F. Wilson, D. Toniolo, P. Gasparini, N. Soranzo, M. S. Sandhu, and J. Marchini. A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet.*, 10(4):e1004234, Apr 2014.

[4] Guido Schwarzer. *meta: Meta-Analysis with R*, 2014. URL `http://CRAN.R-project.org/package=meta`. R package version 4.0-1.

[5] Z. Su, J. Marchini, and P. Donnelly. HAPGEN2: simulation of multiple disease SNPs. *Bioinformatics*, 27(16):2304–2305, Aug 2011.

[6] J. Yang, B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, and P. M. Visscher. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.*, 42(7): 565–569, Jul 2010.

[7] J. Yang, S. H. Lee, M. E. Goddard, and P. M. Visscher. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, 88(1):76–82, Jan 2011.